The Clustering of Galaxies in the Completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmic Flows and Cosmic Web from Luminous Red Galaxies

Metin Ata[1]⋆, Francisco-Shu Kitaura[1,2,3,4,5]†, Chia-Hsun Chuang[6,1],
Sergio Rodríguez-Torres[6,7,8], Raul E. Angulo[9], Simone Ferraro[2,3], Hector Gil-Marín[10,11,12],
Patrick McDonald[2], Carlos Hernández Monteagudo[9], Volker Müller[1], Gustavo Yepes[8],
Mathieu Autefage[1], Falk Baumgarten[1], Florian Beutler[12], Joel R. Brownstein[13],
Angela Burden[14], Daniel J. Eisenstein[15], Hong Guo[16], Shirley Ho[17], Cameron McBride[15],
Mark Neyrinck[18], Matthew D. Olmstead[19], Nikhil Padmanabhan[14], Will J. Percival[12],
Francisco Prada[6,7,20], Graziano Rossi[21], Ariel G. Sánchez[22], David Schlegel[3],
Donald P. Schneider[23,24], Hee-Jong Seo[25], Alina Streblyanska[4], Jeremy Tinker[26],
Rita Tojeiro[27], Mariana Vargas-Magana[28]
Affiliations are listed at the end of the paper

18 January 2017.

ABSTRACT

We present a Bayesian phase-space reconstruction of the cosmic large-scale matter density and velocity fields from the SDSS-III Baryon Oscillations Spectroscopic Survey Data Release 12 (BOSS DR12) CMASS galaxy clustering catalogue. We rely on a given ΛCDM cosmology, a mesh resolution in the range of 6-10 h^{-1} Mpc, and a lognormal-Poisson model with a redshift dependent nonlinear bias. The bias parameters are derived from the data and a general renormalised perturbation theory approach. We use combined Gibbs and Hamiltonian sampling, implemented in the ARGO code, to iteratively reconstruct the dark matter density field and the coherent peculiar velocities of individual galaxies, correcting hereby for coherent redshift space distortions (RSD). Our tests relying on accurate N-body based mock galaxy catalogues, show unbiased real space power spectra of the nonlinear density field up to $k \sim 0.2\,h$ Mpc^{-1}, and vanishing quadrupoles down to $r \sim 20\,h^{-1}$ Mpc. We also demonstrate that the nonlinear cosmic web can be obtained from the tidal field tensor based on the Gaussian component of the reconstructed density field. We find that the reconstructed velocities have a statistical correlation coefficient compared to the true velocities of each individual lightcone mock galaxy of $r \sim 0.68$ including about 10% of satellite galaxies with virial motions (about $r = 0.75$ without satellites). The power spectra of the velocity divergence agree well with theoretical predictions up to $k \sim 0.2\,h$ Mpc^{-1}. This work will be especially useful to improve, e.g. BAO reconstructions, kinematic Sunyaev-Zeldovich (kSZ), integrated Sachs-Wolfe (ISW) measurements, or environmental studies.

Key words: cosmology: theory – large-scale structure of the Universe – catalogues – galaxies: statistics – methods: numerical

1 INTRODUCTION

The large-scale structure of the Universe is a key observable probe to study cosmology. Galaxy redshift surveys provide a three dimensional picture of the distribution of luminous

⋆ E-mail:mata@aip.de
† E-mail:kitaura@aip.de

© 0000 The Authors

tracers across the history of the Universe after cosmic dawn. The recovery of this information relies on accurate modelling of effects including the survey geometry, radial selection functions, galaxy bias, and redshift space distortions caused by the peculiar motions of galaxies.

Many studies require reliable reconstructions of the large-scale gravitational potential from which also the coherent peculiar velocities can be derived. This is the case of the integrated Sachs Wolfe effect (see e.g., Granett et al. 2008; Ilić et al. 2013), the kinematic Sunyaev-Zeldovich effect (see e.g., Planck Collaboration et al. 2016; Hernández-Monteagudo et al. 2015; Schaan et al. 2015), the cosmic flows (e.g. Watkins et al. 2009; Lavaux et al. 2010; Branchini et al. 2012; Courtois et al. 2012; Kitaura et al. 2012c; Heß & Kitaura 2016), or the baryon acoustic oscillations (BAO) reconstructions (see e.g. Eisenstein et al. 2007; Padmanabhan et al. 2012; Anderson et al. 2014a; Ross et al. 2015). Also environmental studies of galaxies demonstrated to benefit from accurate density and velocity reconstructions (see Nuza et al. 2014).

In addition, a number of works have suggested nonlinear transformations, Gaussianising the density field to obtain improved cosmological constraints (Neyrinck et al. 2009, 2011; Yu et al. 2011; Joachimi et al. 2011; Carron & Szapudi 2014; Simpson et al. 2016). Also, linearised density fields can yield improved displacement and peculiar velocity fields (Kitaura & Angulo 2012; Kitaura et al. 2012b; Falck et al. 2012).

Nevertheless, all these studies are affected by redshift space distortions and the sparsity of the signal, which must be handled carefully (McCullagh et al. 2016). Indeed, Seljak (2012) has pointed out that if not properly modeled, nonlinear transformations on density fields including redshift space distortions can lead to biased results. Such a careful modeling is one motivation for the current work.

The inferred galaxy line-of-sight position is a combination of the so-called Hubble flow, i.e. their real distance, and their peculiar motion. The modifications produced by this effect are referred to as redshift space distortions (RSD). They can be used to constrain the nature of gravity and cosmological parameters (see e.g. Berlind et al. 2001; Zhang et al. 2007; Jain & Zhang 2008; Guzzo et al. 2008; Nesseris & Perivolaropoulos 2008; Song & Koyama 2009; Song & Percival 2009; Percival & White 2009; McDonald & Seljak 2009; White et al. 2009; Song et al. 2011; Zhao et al. 2010; Song et al. 2010, for recent studies). The measurement of RSD have in fact become a common technique (Cole et al. 1995; Peacock et al. 2001; Percival et al. 2004; da Ângela et al. 2008; Okumura et al. 2008; Guzzo et al. 2008; Blake et al. 2011; Jennings et al. 2011; Kwan et al. 2012; Samushia et al. 2012; Reid et al. 2012; Okumura et al. 2012; Chuang & Wang 2013a,b; Chuang et al. 2013b,a; Samushia et al. 2013; Zheng et al. 2013; Blake et al. 2013; de la Torre et al. 2013; Beutler et al. 2014; Samushia et al. 2014; Sánchez et al. 2014; Bel et al. 2014; Tojeiro et al. 2014; Okumura et al. 2014; Beutler et al. 2014; Wang 2014; Alam et al. 2015b). These studies are usually based on the large-scale anisotropic clustering displayed by the galaxy distribution in redshift space, although N-body based models for fitting the data to smaller scales have been presented in Reid et al. (2014); Guo et al. (2015a,b, 2016). A recent study suggested to measure the growth rate from density reconstructions (Granett et al.

2015). However, instead of correcting redshift space distortions, these were included in the power spectrum used to recover the density field in redshift space.

Different approaches have been proposed in the literature to recover the peculiar velocity field from galaxy distributions (Yahil et al. 1991; Gramann 1993; Zaroubi et al. 1995; Fisher et al. 1995; Davis et al. 1996; Croft & Gaztanaga 1997; Monaco & Efstathiou 1999; Branchini et al. 2002; Lavaux et al. 2008; Branchini et al. 2012; Wang et al. 2012; Kitaura et al. 2012c; Heß & Kitaura 2016), based on various density-velocity relations (see Nusser et al. 1991; Bernardeau 1992; Chodorowski et al. 1998; Bernardeau et al. 1999; Kudlicki et al. 2000; Mohayaee & Tully 2005; Bilicki & Chodorowski 2008; Jennings & Jennings 2015; Kitaura et al. 2012b; Nadkarni-Ghosh & Singhal 2016).

The main objective of this paper is to perform a self-consistent inference analysis of the density and peculiar velocity field on large scales accounting for all the above mentioned systematic effects (survey geometry, radial selection function, galaxy bias, RSD, non-Gaussian statistics, shot noise). We will rely on the lognormal-Poisson model within the Bayesian framework (Kitaura et al. 2010) to infer the density field from the galaxy distribution. Lognormal-Poisson Bayesian inference performed independently on each density cell reduces to a sufficient statistic characterizing the density field at the two-point level (Carron & Szapudi 2014), but including the density covariance matrix as done here carries additional statistical power. Furthermore we will iteratively solve for redshift space distortions relying on linear theory (Kitaura et al. 2016b).

More complex priors describing the density field than the lognormal assumption can be used (Coles & Jones 1991), based on perturbation theory (Kitaura 2013; Jasche & Wandelt 2013; Wang et al. 2013; Heß et al. 2013), or even on particle mesh approaches (Wang et al. 2014). Also the likelihood describing the statistical distribution of galaxies can be improved modelling the deviation from Poissonity (Ata et al. 2015). Moreover, the relation between the density and the peculiar velocity field could be more accurately modelled including tidal field tensors (Kitaura et al. 2012b). In this work, we want to focus however, on the simplest and most efficient models, which permit us to make the least assumptions with the smallest number of parameters. We leave a more complex nonlinear analysis for future work. In fact we will see that with simple models we can recover the large-scale density and peculiar velocity in the presence of light-cone, survey mask, and selection function effects, with a given ΛCDM cosmology, having chosen the resolution at which our models apply (6-10 h^{-1} Mpc). The majority of previous Bayesian density field reconstructions applied to galaxy redshift surveys did not correct for the anisotropic redshift space distortions (see e.g. Erdogdu et al. 2004; Kitaura et al. 2009; Jasche et al. 2010, 2015; Granett et al. 2015). We aim at filling that gap in this work, and think that the approach presented in this work could become standard in the analysis of galaxy surveys due to its efficiency, simplicity, and its critical accuracy isotropizing the galaxy distribution while dealing with survey masks, selection functions and bias.

Ongoing and future surveys, such as the BOSS[1] (White et al. 2011; Bolton et al. 2012; Alam et al. 2015a), eBOSS (Dawson et al. 2013), DESI[2]/BigBOSS (Schlegel et al. 2011), DES[3] (Frieman & Dark Energy Survey Collaboration 2013), LSST [4] (LSST Dark Energy Science Collaboration 2012), J-PAS[5] (Benitez et al. 2014), 4MOST[6] (de Jong et al. 2012) or Euclid[7] (Laureijs 2009), will require special data analysis techniques, like the one presented here, to extract the maximum available cosmological information.

The structure of this paper is as follows: in Sec. 2 we present the main aspects of our reconstruction method and the ARGO-code (**A**lgorithm for **R**econstructing the **G**alaxy traced **O**verdensities). We emphasize the challenges of dealing with a galaxy redshift survey including cosmic evolution and the novel improvements to this work. In Sec. 3 we describe the BOSS CMASS DR12 data and the mock galaxy catalogues used in this study. In Sec. 4 we show and evaluate the results of our application. We finally present the conclusions in Sec. 5.

2   METHOD

Our basic approach relies on an iterative Gibbs-sampling method, as proposed in Kitaura & Enßlin (2008); Kitaura et al. (2012a) and presented in more detail in Kitaura et al. (2016b). The first step samples linear density fields defined on a mesh δ_L with N_c cells compatible with the number counts on that mesh N_G of the galaxy distribution in real space $\{r\}$. The second step obtains the real space distribution for each galaxy given its observed redshift space s^{obs} position required for the first step, from sampling the peculiar velocities $\{v(\delta_L, f_\Omega)\}$ (with the growth rate given by $f_\Omega \equiv d\log D(a)/d\log a$, and $D(a)$ being the growth factor for a scale factor $a = 1/(1+z)$ or redshift z), assuming that the density field and the growth rate f_Ω are known. The Gibbs-sampling conditional probablity distribution functions can be written as follows showing the quantities, linear densities δ_L and a set of galaxies in real space $\{r\}$, which are sampled from the corresponding conditional PDFs:

$$\delta_L \quad \curvearrowleft \quad \mathcal{P}_\delta\left(\delta_L | N_G(\{r\}), w, C_L(\{p_c\}), \{b_p\}\right), \quad (1)$$

$$\{r\} \quad \curvearrowleft \quad \mathcal{P}_r\left(\{r\} | \{s^{obs}\}, \{v(\delta_L, f_\Omega)\}\right), \quad (2)$$

which is equivalent to sample from the following joint probability distribution function:

$$\mathcal{P}_{joint}\left(\delta_L, \{r\} | \{s^{obs}\}, w, C_L(\{p_c\}), \{b_p\}, f_\Omega\right). \quad (3)$$

To account for the angular completeness (survey mask) and radial selection function, we need to compute the 3D completeness w defined on the same mesh, as the density field (see e.g. Kitaura et al. 2009). Also we have to assume a given covariance matrix $C_L \equiv \langle \delta_L^\dagger \delta_L \rangle$ (a $N_c \times N_c$ matrix),

---

[1] http://www.sdss3.org/surveys/boss.php
[2] http://desi.lbl.gov/
[3] http://www.darkenergysurvey.org
[4] http://www.lsst.org/lsst/
[5] http://j-pas.org/
[6] https://www.4most.eu/
[7] http://www.euclid-ec.org

determined by a set of cosmological parameters $\{p_c\}$ within a ΛCDM framework. We aim at recovering the dark matter density field which governs the dynamics of galaxies. Since galaxies are biased tracers, we have to assume some parametrised model relating the density field to the galaxy density field with a set of bias parameters $\{b_p\}$. We note that assuming a wrong growth rate will yield an anisotropic reconstructed density field. A recent work investigated this by jointly sampling the anisotrpic power spectrum including the growth rate and the redshift space density field (see Granett et al. 2015).

After these probabilities reach their so-called stationary distribution, the drawn samples are represetatives of the target distribution. In the following we define Eqs. 1 and 2 in detail and describe our sampling strategy.

2.1   Density sampling

The posterior probability distribution of Eq. 1 is sampled using a Hamiltonian Monte Carlo (HMC) technique (see Duane et al. 1987). For a comprehensive review see Neal (2012). This technique has been applied in cosmology in a number of works (see e.g. Taylor et al. 2008; Jasche & Kitaura 2010; Jasche et al. 2010; Kitaura et al. 2012a,c; Kitaura 2013; Wang et al. 2013, 2014; Ata et al. 2015; Jasche & Wandelt 2013). To apply this technique to our Bayesian reconstruction model, we need to define the posterior distribution function through the product of a prior π (see Sec. 2.1.1) and a likelihood $\mathcal{L}$ (see Sec. 2.1.2) which up to a normalisation is given by

$$\mathcal{P}_\delta\left(\delta_L | N_G(\{r\}), w, C(\{p_c\}), \{b_p\}\right) \propto \quad (4)$$

$$\pi(\delta_L | C(\{p_c\})) \times \mathcal{L}(N_G | \rho_G^{obs}, \{b_p\}), \quad (5)$$

with ρ_G^{obs} being the expected number counts per volume element. The overall sampling strategy then is enclosed in Sec. 2.1.3.

2.1.1   Lognormal Prior

As a prior we rely on the lognormal structure formation model introduced in Coles & Jones (1991). This model gives an accurate description of the matter statistics (of the cosmic evolved density contrast $\delta \equiv \rho/\bar{\rho} - 1$) on scales larger than about 6-10 h^{-1} Mpc (see e.g. Kitaura et al. 2009). In such a model one considers that the logarithmically transformed density field δ_L is a good representation of the linear density field

$$\delta_L \equiv \log(1 + \delta) - \mu, \quad (6)$$

with

$$\mu \equiv \langle \log(1 + \delta) \rangle, \quad (7)$$

and is Gaussian distributed with zero mean and a given covariance matrix C_L

$$-\ln \pi(\delta_L | C_L(\{p_c\})) = \frac{1}{2} \delta_L^\dagger C_L^{-1} \delta_L + c, \quad (8)$$

with c being some normalisation constant of the prior. This model yields, however, a poor description of the three-point statistics (see White et al. 2014; Chuang et al. 2015), and will have a different mean field μ depending on the higher order statistics of the dark matter field. The mean field computed

based on the density field, as obtained from N-body simulations using the definition in Eq. 7, can strongly deviate from the theoretical prediction for lognormal fields $\mu = -\sigma^2/2$ depending on the resolution (with σ^2 being the variance of the field $\boldsymbol{\delta}_{\rm L}$). In fact, if one expands the logarithm of the density field in a series with the first term being the linear density field followed by all the higher order terms $\boldsymbol{\delta}^+$ (see Kitaura & Angulo 2012)

$$\log(1 + \boldsymbol{\delta}) = \boldsymbol{\delta}_{\rm L} + \boldsymbol{\delta}^+ \,, \qquad (9)$$

one finds that the mean field depends on the order of the expansion

$$\mu \equiv \langle \log(1 + \boldsymbol{\delta}) \rangle = \langle \boldsymbol{\delta}^+ \rangle \,. \qquad (10)$$

In practice, the data will determine the mean field μ. In unobserved regions, the mean field should be given by the theoretical lognormal value ($\mu = -\sigma^2/2$). In observed regions, the number density and completeness will determine the value of the mean field. Since galaxy redshift surveys have in general a varying completeness as a function of distance, the assumption of a unique mean field can introduce an artificial radial selection function. For this reason we suggest to follow Kitaura et al. (2012a) and iteratively sample the mean field from the reconstructed linear density field assuming large enough volumes $\langle \delta \rangle = 0 = \langle e^{\delta_{\rm L} + \mu} - 1 \rangle$, i.e., $\mu = -\ln(\langle e^{\delta_{\rm L}} \rangle)$. The assumption that volume averages of the linear and nonlinear density field vanish in the ensemble average, does not imply that this happens for the individual reconstructions, which will be drawn from our posterior analysis allowing for cosmic variance. We will consider, as a crucial novel contribution, individual redshift z and completeness w bins

$$\mu_{(z,w)} = -\ln(\langle e^{\delta_{\rm L}} \rangle_{(z,w)}) \,. \qquad (11)$$

This can be expressed as an additional Gibbs-sampling step

$$\mu_{(z,w)} \quad \curvearrowleft \quad \mathcal{P}_\mu \left(\mu_{(z,w)} | \boldsymbol{\delta}_{\rm L}(\boldsymbol{r}, z), \boldsymbol{w} \right) \,. \qquad (12)$$

In this way we account for redshift and completeness dependent renormalised lognormal priors. In practice, since the evolution of the three-point statistics can be considered to be negligible within the covered redshift range for CMASS galaxies (Kitaura et al. 2016a), we will perform the ensemble average only in completeness bins.

2.1.2   *Likelihood and data model*

The likelihood describes the data model. In our case the probability to draw a particular number of galaxy counts $N_{{\rm G}i}$ per cell i, given an expected number count per cell $\rho_{{\rm G}i}^{\rm obs}$, is modelled by the Poisson distribution function

$$-\ln \mathcal{L}(\boldsymbol{N}_{\rm G} | \boldsymbol{\rho}_{\rm G}^{\rm obs}, \{b_{\rm p}\}) = \sum_i^{Nc} \left(-N_{{\rm G}i} \ln \rho_{{\rm G}i}^{\rm obs} + \rho_{{\rm G}i}^{\rm obs} \right) + c, \quad (13)$$

with $N_{\rm c}$ being the total number of cells of the mesh, and c being some normalisation constant of the likelihood. This expectation value is connected to the underlying matter density δ_i by the particular chosen bias model $\mathcal{B}(\boldsymbol{\rho}_{\rm G} | \boldsymbol{\delta})$. In particular, we rely on a power-law bias (linear in the log-density field) connecting the galaxy density field to the underlying dark matter density $\rho_{\rm G} \propto (1 + \delta)^b$ (de la Torre & Peacock 2013). More complex biasing models can be found in the

literature (Fry & Gaztanaga 1993; Cen & Ostriker 1993; McDonald & Roy 2009; Kitaura et al. 2014; Neyrinck et al. 2014; Ahn et al. 2015). In fact threshold bias can be very relevant to describe the three-point statistics of the galaxy field (Kitaura et al. 2015, 2016a), and stochastic bias (Kitaura et al. 2014) is crucial to properly describe the clustering on small scales. All these bias components have been investigated within a Bayesian framework in Ata et al. (2015). We will, however, focus in this work on the two-point statistics on large scales ($k \lesssim 0.2\,h\ {\rm Mpc}^{-1}$), and neglect such deviations. The bias model needs to account for cosmic evolution. In linear theory and within ΛCDM this is described by the growth factor:

$$D(z) = \frac{H(z)}{H_0} \int_z^\infty dz' \frac{(1+z')}{H^3(z')} \bigg/ \int_0^\infty dz' \frac{(1+z')}{H^3(z')} \,, \qquad (14)$$

permitting one to relate the density field at a given redshift to a reference redshift $z_{\rm ref}$: $\delta_i(z_{\rm ref}) = G(z_{\rm ref}, z_i)\,\delta_i(z_i)$ with

$$G(z_{\rm ref}, z_i) \equiv D(z_i)/D(z_{\rm ref}) \,. \qquad (15)$$

The reference redshift must be chosen to be lower than the lowest redshift in the considered volume to ensure that the growth factor ratio $G(z_{\rm ref}, z_i) \equiv D(z_i)/D(z_{\rm ref})$ remains below one. Otherwise, negative densities will arise in low density cells, causing singularities in the lognormal model. Another important ingredient in our model is the angular mask and radial selection function describing the three dimensional completeness $\boldsymbol{w}$, which can be seen as a response function between the signal and the data: $\rho_{{\rm G}i}^{\rm obs} \equiv w_i \rho_{{\rm G}i} \propto w_i \mathcal{B}(\boldsymbol{\rho}_{\rm G}|\boldsymbol{\delta})|_i$ (see e.g. Kitaura & Enßlin 2008). One needs to consider now, that only when the bias is linear the proportionality factor is given by the mean number density $\bar{N} \equiv \langle \boldsymbol{\rho}_{\rm G} \rangle$: $\rho_{{\rm G}i} = \bar{N}(1 + b_{\rm L}\delta)$, with $b_{\rm L}$ being the linear bias. This model is inconvenient for bias larger than one, as it is the case of luminous red galaxies, since negative densities could arise. In the general case, the proportionality constant will be given by the bias model (Kitaura et al. 2014)

$$\gamma(z) \equiv \bar{N} / \langle \mathcal{B}(\boldsymbol{\rho}_{\rm G}|\boldsymbol{\delta}) \rangle_{(z)} \,, \qquad (16)$$

which we suggest to iteratively sample from the reconstructed density field in redshift bins. If we instead use a model defined as

$$\rho_{{\rm G}i} \equiv \bar{N} \left(1 + \mathcal{B}(\boldsymbol{\rho}_{\rm G}|\boldsymbol{\delta})|_i - \langle \mathcal{B}(\boldsymbol{\rho}_{\rm G}|\boldsymbol{\delta}) \rangle \right) \,,$$

which also ensures the correct mean number density by construction, negative expected number counts are allowed, which we want to avoid. For this reason we will rely on the following bias model:

$$\rho_{{\rm G}i}^{\rm obs} \equiv w_i \gamma(z_i)(1 + G(z_i, z_{\rm ref})\delta_i)^{b_{\rm L}(z_i)f_{\rm b}} \,, \qquad (17)$$

where we have included a bias correction factor $f_{\rm b}$, which accounts for the deviation between linear and power-law bias. With this model, the sampling of the normalisation constant can be expressed as an additional Gibbs-sampling step

$$\gamma_{(z)} \quad \curvearrowleft \quad \mathcal{P}_\gamma \left(\gamma_{(z)} | \bar{N}, \boldsymbol{\delta}, G(z, z_{\rm ref}), b_{\rm L}(z), f_{\rm b} \right) \,. \qquad (18)$$

Given a redshift z one can define the ratio between the galaxy correlation function in redshift space at z ($\xi_{\rm G}^s(z)$) and the matter correlation function in real space at $z_{\rm ref}$ ($\xi_{\rm M}(z_{\rm ref})$) as

$$c_{\rm L}^s(z) \equiv \sqrt{\xi_{\rm G}^s(z)/\xi_{\rm M}(z_{\rm ref})} \,. \qquad (19)$$

The quantity $\xi_{\mathrm{G}}^s(z)$ can be obtained from the data without having to assume any bias, nor growth rate. Furthermore, one can use the Kaiser factor ($K = 1 + 2/3 f_\Omega/b_{\mathrm{L}} + 1/5(f_\Omega/b_{\mathrm{L}})^2$, with f_Ω being the growth rate, Kaiser 1987) to relate the galaxy correlation function in redshift space to the matter real space correlation function

$$
\begin{aligned}
\xi_{\mathrm{G}}^s(z) &= K(z)\,\xi_{\mathrm{G}}(z) \\
&= K(z)\,b_{\mathrm{L}}^2(z)\,G^2(z, z_{\mathrm{ref}})\,\xi_{\mathrm{M}}(z_{\mathrm{ref}}) \,. \quad (20)
\end{aligned}
$$

From the last two equations we find a quadratic expression for $b_{\mathrm{L}}(z)$ for each redshift z

$$
b_{\mathrm{L}}^2(z) + \frac{2}{3} f_\Omega(z) b_{\mathrm{L}}(z) + \frac{1}{5} f_\Omega^2(z) - \frac{(c_{\mathrm{L}}^s(z))^2}{G^2(z, z_{\mathrm{ref}})} = 0 \,, \quad (21)
$$

with only one positive solution, leaving the bias correction factor f_{b} as a potential free parameter in our model (see the renormalised perturbation theory based derivation below)

$$
b_{\mathrm{L}}(z) = -\frac{1}{3} f_\Omega(z) + \sqrt{-\frac{4}{45} f_\Omega(z)^2 + (c_{\mathrm{L}}^s(z))^2 \left(\frac{D(z_{\mathrm{ref}})}{D(z)}\right)^2} \,. \quad (22)
$$

By coincidence, the bias measured in redshift space on large scales $c_{\mathrm{L}}^s(z) = 1.84 \pm 0.1$ (with respect to the dark matter power spectrum at redshift $z = 0.57$) is constant for CMASS galaxies across the considered redshift range (see, e.g., Rodríguez-Torres et al. 2015). Nevertheless, the (real space) linear bias $b_{\mathrm{L}}(z)$ is not, as it needs to precisely compensate for the growth of structures (growth factor) and the evolving growth rates, ranging between 2.00 and 2.30. The nonlinear bias correction factor f_{b} is expected to be less than "one", since we are using the linear bias in the power-law. One can predict f_{b} from renormalised perturbation theory, which in general, will be a function of redshift. Let us Taylor expand our bias expression (Eq. 17) to third order

$$
\delta_{\mathrm{g}}(z_i) \equiv \frac{\rho_{\mathrm{g}}}{\bar{\rho}_{\mathrm{g}}}(z_i) - 1 \simeq b_{\mathrm{L}}(z_i) f_{\mathrm{b}}(z_i)\delta(z_i) \quad (23)
$$

$$
+\frac{1}{2} b_{\mathrm{L}}(z_i) f_{\mathrm{b}}(z_i)(b_{\mathrm{L}}(z_i) f_{\mathrm{b}}(z_i) - 1)\left((\delta(z_i))^2 - \sigma^2(z_i)\right) +
$$

$$
\frac{1}{3!} b_{\mathrm{L}}(z_i) f_{\mathrm{b}}(z_i)(b_{\mathrm{L}}(z_i) f_{\mathrm{b}}(z_i) - 1)(b_{\mathrm{L}}(z_i) f_{\mathrm{b}}(z_i) - 2)\,(\delta(z_i))^3 \,,
$$

with $\delta(z_i) = G(z_i, z_{\mathrm{ref}})\delta(z_{\mathrm{ref}})$. The usual expression for the perturbatively expanded overdensity field to third order ignoring nonlocal terms is given by

$$
\delta_{\mathrm{g}}(z_i) = c_\delta(z_i)\delta(z_i) + \frac{1}{2} c_{\delta^2}(z_i)(\delta^2(z_i) - \sigma^2(z_i)) + \frac{1}{3!} c_{\delta^3}(z_i)\delta^3(z_i) \,. \quad (24)
$$

Correspondingly, one can show that the observed, renormalised, linear bias is given by (see McDonald & Roy 2009)

$$
b_\delta(z_i) = c_\delta(z_i) + \frac{34}{21} c_{\delta^2}(z_i)\sigma^2(z_i) + \frac{1}{2} c_{\delta^3}(z_i)\sigma^2(z_i) \,. \quad (25)
$$

By considering that in our case the observable linear bias is expected to be given by $b_{\mathrm{L}}(z_i)$ and identifying the coefficients $\{c_\delta = f_{\mathrm{b}} b_{\mathrm{L}}, \; c_{\delta^2} = f_{\mathrm{b}} b_{\mathrm{L}}(f_{\mathrm{b}} b_{\mathrm{L}} - 1), \; c_{\delta^3} = f_{\mathrm{b}} b_{\mathrm{L}}(f_{\mathrm{b}} b_{\mathrm{L}} - 1)(f_{\mathrm{b}} b_{\mathrm{L}} - 2)\}$ from Eqs. 23 and 24 one can derive the following cubic equation for f_{b}

$$
f_{\mathrm{b}}^3 \left(\frac{1}{2} b_{\mathrm{L}}^3(z_i)\sigma^2(z_i)\right) \quad (26)
$$

$$
+f_{\mathrm{b}}^2 \left(-\frac{3}{2} b_{\mathrm{L}}^2(z_i)\sigma^2(z_i) + \frac{34}{21} b_{\mathrm{L}}^2(z_i)\sigma^2(z_i)\right)
$$

$$
+f_{\mathrm{b}}\, b_{\mathrm{L}}(z_i) \left(1 + \left(-\frac{34}{21} + 1\right)\sigma^2(z_i)\right) - b_{\mathrm{L}}(z_i) = 0 \,.
$$

Let us consider the case of a cell resolution of $6.25\,h^{-1}\,\mathrm{Mpc}$. The only real solutions for redshift $z = 0.57$ ($G = 0.78$) and $b_{\mathrm{L}} = 2.1 \pm 0.1$, are $f_{\mathrm{b}} = 0.62 \pm 0.01$ including the variance from the nonlinear transformed field ($\sigma^2(\delta) = 1.75$), and $f_{\mathrm{b}} = 0.71 \pm 0.02$ including the variance from the linear field ($\sigma^2(\delta_{\mathrm{L}}) = 0.91$). This gives us a hint of the uncertainty in the nonlinear expansion. Let us, hence, quote as the theoretical prediction for the bias correction factor the average between both mean values with the uncertainty given by the difference between them $f_{\mathrm{b}} = 0.66 \pm 0.1$. These results show little variation (± 0.01) across the redshift range (see §3.2). Leaving f_{b} as a free parameter and sampling it to match the power spectrum on large scales yields $f_{\mathrm{b}} = 0.7 \pm 0.05$ (see §4). Although there is an additional uncertainty associated to this measure, since the result depends on the particular k mode range used in the goodness of fit. Therefore, one can conclude that the theoretical predictions account for the nonlinear correction within the associated uncertainties on large scales in terms of the two point statistics. We include only delta bias terms in Eqs. 24, 25, because these equations describe the model we implemented, represented by Eqs. 17, 23, where we did not include any tidal bias. As shown by McDonald & Roy (2009), the only effect of tidal bias terms in the low k (large scale) limit is to renormalize the standard linear delta bias (and shot noise). We are therefore implicitly including these effects if present in the data when we fit for the bias (Eq. 22), i.e., our model is complete in the low-k limit. As we go to higher k, i.e., smaller scales, tidal bias can have a non-trivial effect in the model (McDonald & Roy 2009), along with various other non-linear effects which enter at the same order in perturbation theory (i.e., non-linear gravitational evolution, higher order density bias different from that implied by Eq. 17, non-linearity/biases related to the redshift space transformation). These effects could be included in future models for higher accuracy.

2.1.3 Hamiltonian Monte Carlo of the linear density field

In this section we recap the Hamiltonian Monte Carlo sampling technique (HMC) to sample the matter density within the Bayesian framework. This technique requires the gradients of the lognormal-Poisson model, as introduced in Kitaura et al. (2010). The HMC technique was first applied to this model with a linear bias in Jasche & Kitaura (2010) and later with more complex bias relations and likelihoods in Ata et al. (2015). In this approach one defines a potential energy $U(\boldsymbol{x})$, given by the negative logarithm of the posterior distribution function, and a kinetic energy $\boldsymbol{K}(\boldsymbol{p})$

$$
\begin{aligned}
U(\boldsymbol{x}) &= -\ln \mathcal{P}(\boldsymbol{x}) \quad (27) \\
\mathcal{H}(\boldsymbol{x}, \boldsymbol{p}) &= U(\boldsymbol{x}) + K(\boldsymbol{p}) \,, \quad (28)
\end{aligned}
$$

where the Hamiltonian $\mathcal{H}(\boldsymbol{x}, \boldsymbol{p})$ is given by the sum of the potential and the kinetic energy. In this formalism we use $\boldsymbol{x}$ as a pseudo spatial variable (in our case the linear density field $\boldsymbol{\delta}_{\mathrm{L}}$) and $\boldsymbol{p}$ as the conjugate momentum. HMC requires the computation of the negative logarithm of Eq. 5 and its derivatives with respect to the sampled quantity (the linear density field $\boldsymbol{\delta}_{\mathrm{L}}$ in our case). The *kinetic energy* term is constructed on the nuisance parameters given by the *momenta*

p and *mass* variance $\mathbf{M}$:

$$K(\boldsymbol{p}) \equiv \frac{1}{2} \sum_{ij} p_i M_{ij}^{-1} p_j \,. \tag{29}$$

The *canonical* distribution function defined by the Hamiltonian (or the joint distribution function of the signal and *momenta*) is then given by:

$$\begin{aligned} P(\boldsymbol{x}, \boldsymbol{p}) &= \frac{1}{Z_H} \exp(-\mathcal{H}(\boldsymbol{s}, \boldsymbol{p})) \\ &= \left[\frac{1}{Z_K} \exp(-K(\boldsymbol{p}))\right] \left[\frac{1}{Z_E} \exp(-U(\boldsymbol{x}))\right] \\ &= P(\boldsymbol{p}) P(\boldsymbol{x}) \,, \end{aligned} \tag{30}$$

with Z_H, Z_K and Z_E being the *partition* functions so that the probability distribution functions are normalised to one. In particular, the normalisation of the Gaussian distribution for the *momenta* is represented by the *kinetic partition* function Z_K. The Hamiltonian sampling technique does not require the terms which are independent of the configuration coordinates as we will show below.

From Eq. (30) it can be noticed that in case we have a method to sample from the joint distribution function $P(\boldsymbol{x}, \boldsymbol{p})$, marginalizing over the momenta we can in fact, sample the posterior $P(\boldsymbol{x})$.

The Hamiltonian dynamics provides such a method. We can define a dynamics on *phase-space* (positions and momenta) with the introduction of a *time* parameter t. The Hamiltonian equations of motion are given by:

$$\frac{dx_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i} = \sum_j M_{ij}^{-1} p_j \,, \tag{31}$$

$$\frac{dp_i}{dt} = -\frac{\partial \mathcal{H}}{\partial x_i} = -\frac{\partial U(x)}{\partial x_i} \,. \tag{32}$$

To sample the posterior one has to solve these equations for randomly drawn *momenta* according to the kinetic term defined by Eq. (29). This is done by drawing Gaussian samples with a variance given by the *mass* $\mathbf{M}$ which can tune the efficiency of the sampler (see Jasche & Kitaura 2010). We rely on the Fourier formulation to capture the correlation function through the power spectrum and include some preconditioning diagonal matrices to speed up the algorithm. The marginalization over the *momenta* occurs by drawing new *momenta* for each Hamiltonian step disregarding the ones of the previous step.

It is not possible to follow the dynamics exactly, as one has to use a discretized version of the equations of motion. It is convenient to use the *leapfrog* scheme which has the properties of being *time*-reversible and conserve *phase-space* volume being necessary conditions to ensure *ergodicity*:

$$p_i\left(t + \frac{\epsilon}{2}\right) = p_i(t) - \frac{\epsilon}{2} \left.\frac{\partial U(\boldsymbol{x})}{\partial x_l}\right|_{x_i(t)} \,, \tag{33}$$

$$x_i(t + \epsilon) = x_i(t) + \epsilon \sum_j M_{ij}^{-1} p_j\left(t + \frac{\epsilon}{2}\right) \,, \tag{34}$$

$$p_i(t + \epsilon) = p_i\left(t + \frac{\epsilon}{2}\right) - \frac{\epsilon}{2} \left.\frac{\partial U(\boldsymbol{x})}{\partial x_l}\right|_{x_i(t+\epsilon)} \,. \tag{35}$$

The dynamics of this system are followed for a period of *time* $\Delta\tau$, with a value of ϵ small enough to give acceptable errors and for $N_\tau = \Delta\tau/\epsilon$ iterations. In practice ϵ and N_τ are randomly drawn from a uniform distribution to avoid resonant trajectories (see Neal 1993).

The solution of the equations of motion will move the system from an initial state $(\boldsymbol{s}, \boldsymbol{p})$ to a final state $(\boldsymbol{s}', \boldsymbol{p}')$ after each sampling step. Although the Hamiltonian equations of motion are energy conserving, our approximate solution is not. Moreover, the starting guess will not be drawn from the correct distribution and a *burn-in* phase will be needed. For these reasons a Metropolis-Hastings acceptance step has to be introduced in which the new *phase-space* state $(\boldsymbol{x}', \boldsymbol{p}')$ is accepted with probability:

$$P_A = \min\left[1, \exp(-\delta\mathcal{H})\right] \,, \tag{36}$$

with $\delta\mathcal{H} \equiv \mathcal{H}(\boldsymbol{x}', \boldsymbol{p}') - \mathcal{H}(\boldsymbol{x}, \boldsymbol{p})$.

In particular, the required lognormal-Poisson gradients for the prior and likelihood including cosmic evolution are given by

$$-\frac{\partial}{\partial \boldsymbol{\delta}_{\mathrm{L}}} \ln \pi = \mathbf{C}_{\mathrm{L}}^{-1} \boldsymbol{\delta}_{\mathrm{L}} \,, \tag{37}$$

and

$$-\frac{\partial \ln \mathcal{L}}{\partial \boldsymbol{\delta}_{\mathrm{L}}}\Big|_i = \left(-\frac{N_{\mathrm{G}i}}{\rho_{\mathrm{G}i}^{\mathrm{obs}}} + 1\right) \cdot \frac{b_{\mathrm{L}}(z) f_{\mathrm{b}} \, G(z, z_{\mathrm{ref}})(1 + \delta_i)}{1 + G(z, z_{\mathrm{ref}})\delta_i} \rho_{\mathrm{G}i}^{\mathrm{obs}} \,, \tag{38}$$

respectively. The linear density field is defined at the reference redshift z_{ref}.

2.2 Velocity sampling

The peculiar motions of galaxies can be divided into two categories: coherent flows (Kaiser 1987) and quasi-virialised or dispersed velocities. While the former are well constrained by the large-scale density field, the latter become relevant on smaller nonlinear scales (see e.g. Reid et al. 2014). Thus, one can write the total velocity field as the sum of the curl-free coherent bulk flow, which can directly be inferred from the large-scale density field within linear theory, and the dispersion term $\boldsymbol{v}_{\mathrm{disp}}$

$$\boldsymbol{v}(\boldsymbol{r}, z) = -f_\Omega(a) \, H(a) \, a \, \nabla\nabla^{-2} \delta(\boldsymbol{r}, z) + \boldsymbol{v}_{\mathrm{disp}} \,, \tag{39}$$

where H is the Hubble constant. A simple way of including the dispersion term is to randomly draw it from a Gaussian with a particular standard deviation. One may consider about $50 \, \mathrm{km} \, s^{-1}$ (see §4), the typical 1-σ uncertainty within linear theory (Kitaura et al. 2012b). More precise and sofisticated ways of dealing with quasi-virialised RSD are left for future work (see e.g. Heß et al. 2013; Jennings & Jennings 2015; Kitaura et al. 2016b). Here we aim at focussing on the coherent flows on the limit of vanishing dispersions (see §4 for a comparison study with and w/o dispersion). In practice we are restricting our study to resolutions in the range between 6 and 10 h^{-1} Mpc, which yield robust results on large scales (see Kitaura et al. 2016b). Tidal field corrections could be included in the model (see Kitaura et al. 2012b). Also one could try to get improved velocity reconstructions from the linear component rather than from the nonlinear one as we do here (see Falck et al. 2012; Kitaura & Angulo 2012). Nevertheless, there is a (nearly constant) bias from the lognormal transformation present in the linear density field, which we want to avoid to reduce the number of parameters (see Neyrinck et al. 2009). The mapping between real space and redshift space positions for each individual galaxy is described by

$$\boldsymbol{r}^{j+1} = \boldsymbol{s}^{\mathrm{obs}} - \left(\frac{\boldsymbol{v}\left(\boldsymbol{r}^j, z\right) \cdot \hat{\boldsymbol{r}}}{H(a) \, a}\right) \hat{\boldsymbol{r}} \,, \tag{40}$$
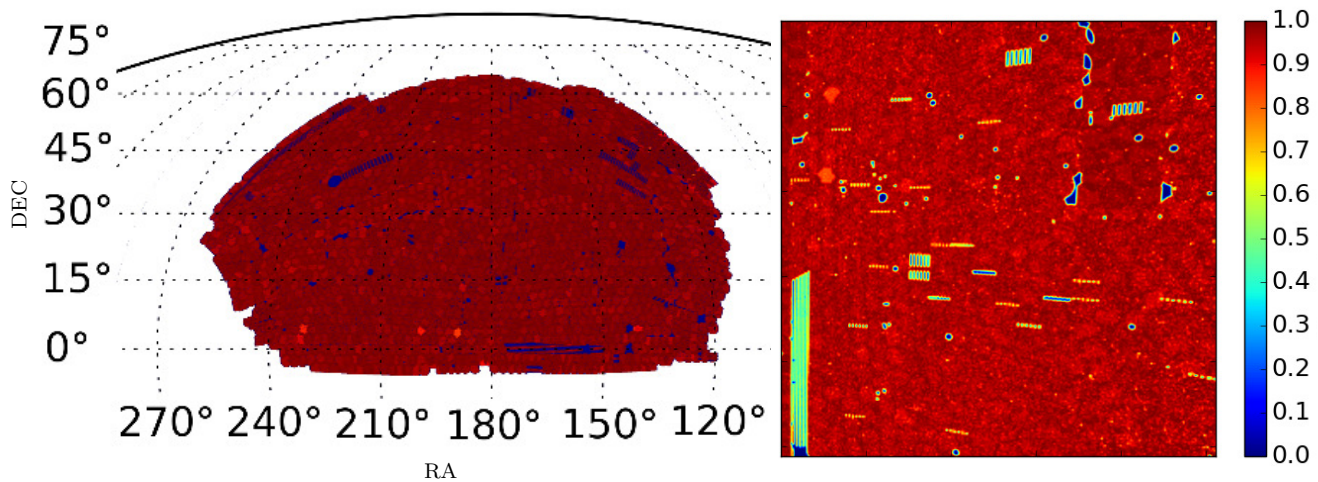
Figure 1. Left panel: Angular mask (right ascension RA vs declination DEC), ranging from zero to one, showing the completeness on the sky of the SDSS-III BOSS DR12 survey. **Right panel:** Slice (in the $x - y$ plane) of the 3D-projected angular mask on a volume of $1250 \, h^{=1}$ Mpc side.
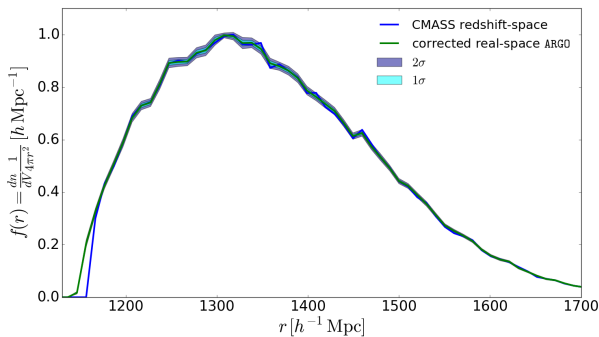

Figure 2. Radial selection function $f(r)$ for a subvolume of the CMASS galaxy survey normalised to unity before and after RSD corrections with ARGO. The mean is calculated by calculating $f(r)$ for 2000 reconstructions.

where j and $j+1$ are two subsequent Gibbs-sampling iterations, and $\hat{r}$ denotes the unit vector in line of sight direction. The peculiar velocity needs to be evaluated in real space, which requires an iterative sampling scheme. Each galaxy requires in principle a peculiar velocity field computed at that redshift, as the growth rate changes with redshift. In practice we construct a number of peculiar velocity fields defined on the same mesh but at different redshifts, i.e., from density fields multiplied with the corresponding growth factors and rates. Each galaxy will get a peculiar velocity field assigned interpolated to its position within the cell taken from the peculiar velocity mesh at the corresponding redshift bin.

3 INPUT DATA

In this paper, we use N-body based mock galaxy catalogues constructed to match the clustering bias, survey mask, selection functions, and number densities of the BOSS DR12 CMASS galaxies. This permits us to test our method, as both real space and redshift space catalogues are known. Finally, we apply our analysis method to the BOSS DR12

CMASS data. Let us describe the input galaxy catalogues below.

3.1 BOSS DR12 galaxy catalogue

This work uses data from the Data Release DR12 (Alam et al. 2015a) of the Baryon Oscillation Spectroscopic Survey (BOSS) (Eisenstein et al. 2011). The BOSS survey uses the SDSS 2.5 meter telescope at Apache Point Observatory (Gunn et al. 2006) and the spectra are obtained using the double-armed BOSS spectrograph (Smee et al. 2013). The data are then reduced using the algorithms described in (Bolton et al. 2012). The target selection of the CMASS and LOWZ samples, together with the algorithms used to create large-scale structure catalogues (the MKSAMPLE code), are presented in Reid et al. (2016).

We restrict this analysis to the CMASS sample of luminous red galaxies (LRGs), which is a complete sample, nearly constant in mass and volume limited between the redshifts $0.43 \leq z \leq 0.7$ (see Anderson et al. (2014b) for details of the targeting strategy).

3.2 Mock galaxy catalogues in real and redshift space

The mock galaxy catalogues used in this study were presented in Rodríguez-Torres et al. (2015), and are extracted from one of the BigMultiDark simulations[8] (Klypin et al. 2016), which was performed using GADGET-2 (Springel et al. 2005) with $3,840^3$ particles on a volume of $(2.5 \, h^{-1} \text{Gpc})^3$ assuming ΛCDM Planck cosmology with $\{\Omega_{\rm M} = 0.307115, \Omega_{\rm b} = 0.048206, \sigma_8 = 0.8288, n_s = 0.9611\}$, and a Hubble constant $(H_0 = 100 \, h \, {\rm km \, s^{-1} \, Mpc^{-1}})$ given by $h = 0.6777$. Haloes were defined based on the Bound Density Maxima (BDM) halo finder (Klypin & Holtzman 1997).

They have been constructed based on the Halo Abundance Matching (HAM) technique to connect haloes to
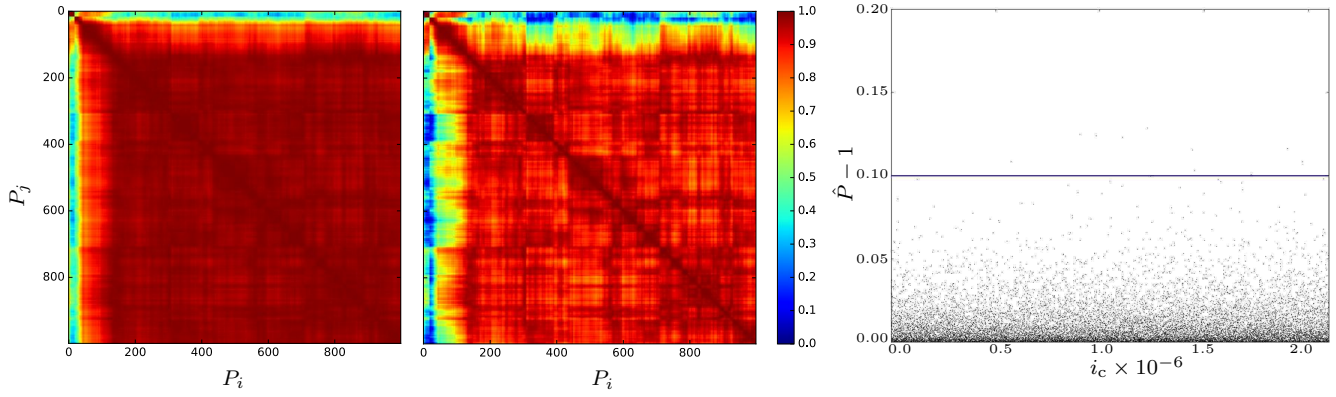
---

[8] http://www.multidark.org/MultiDark/

Figure 3. Convergence analysis of the Gibbs-Hamiltonian sampler. **Left and middle panels:** Power spectrum correlation matrix $\mathcal{R}_{ij}$ of the first 1000 iterations of ARGO with a mesh of 128^3. Each entry of the matrix represents the correlation coefficient of the power spectra P_i and P_j: $\mathcal{R}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}\,C_{jj}}}$, where $C_{ij} = \langle (P_i - \langle P_i \rangle)(P_j - \langle P_j \rangle) \rangle$ is the covariance matrix. **Left panel:** correlation matrix for all modes of the power spectrum, **middle panel:** correlation matrix for the lowest 30 modes, corresponding up to $k = 0.2\,h\,\mathrm{Mpc}^{-1}$. **Right panel:** Potential scale reduction factor $\hat{P}$ of the Gelman & Rubin (1992) test comparing the mean of variances of different chains with the variance of the different chain means. The cell number i_c of a 128^3 mesh is plotted against the potential scale reduction factor $(\hat{P} - 1)$. Commonly a $\hat{P} - 1$ of less then 0.1 (blue line) is required to consider the chains to be converged at the target distribution. Here only two chains were compared, already showing that the majority of cells have converged. This result is already satisfactory, since including more chains will increase the statistics and reduce the potential scale reduction factor, eventually showing that all cells have converged.
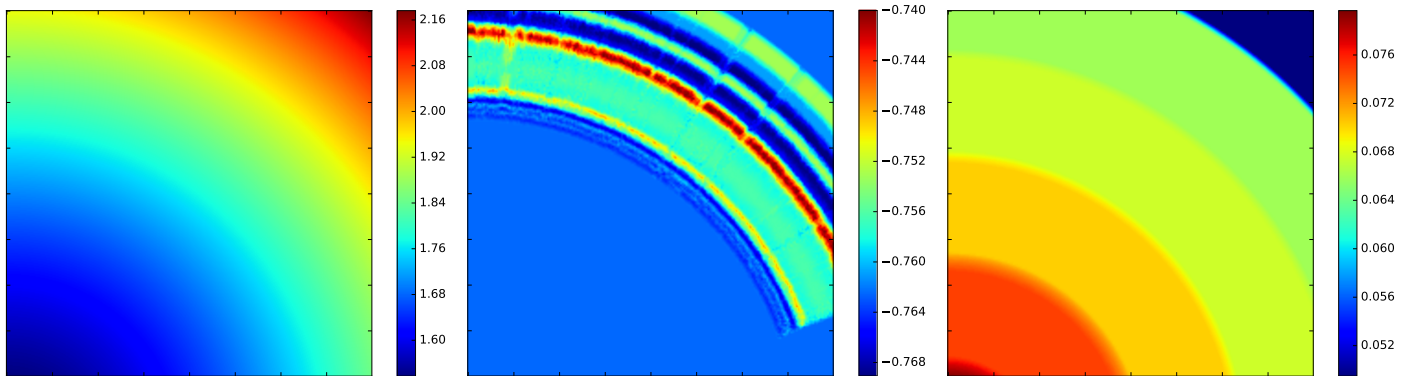
Figure 4. Additional sampled quantities. Based on a light-cone mock catalogue in redshift space with $6.25\,h^{-1}$ Mpc resolution: slices of thickness $\sim 6\,h^{-1}$ Mpc in the $x - z$ plane of the 3D cubical mesh of side $1250\,h^{-1}$ Mpc and 200^3 cells for the following quantities: **left panel:** the linear real space bias b_L multiplied with the nonlinear constant correction factor $f_b = 0.7$, **middle panel:** the lognormal mean field μ, and **right panel:** the galaxy number density normalisation γ.

galaxies (Kravtsov et al. 2004; Neyrinck et al. 2004; Tasitsiomi et al. 2004; Vale & Ostriker 2004; Conroy et al. 2006; Kim et al. 2008; Guo et al. 2010; Wetzel & White 2010; Trujillo-Gomez et al. 2011; Nuza et al. 2013).

At first order HAM assumes a one-to-one correspondence between the luminosity or stellar and dynamical masses: galaxies with more stars are assigned to more massive haloes or subhaloes. The luminosity in a red-band is sometimes used instead of stellar mass. There should be some degree of stochasticity in the relation between stellar and dynamical masses due to deviations in the merger history, angular momentum, halo concentration, and even observational errors (Tasitsiomi et al. 2004; Behroozi et al. 2010; Trujillo-Gomez et al. 2011; Leauthaud et al. 2011). Therefore, we include a scatter in that relation necessary to

accurately fit the clustering of the BOSS data (Rodríguez-Torres et al. 2015).

4 RESULTS

In this section we present the results obtained with the ARGO code including the cosmic evolution treatment described in Sec. 2 on the mock galaxy catalogues and finally on the data. Let us first describe the preparation of the data.

4.1 Preparation of the data

As explained in Sec. 2 our method requires the galaxy number counts on a mesh. Therefore we need first to assume a fiducial cosmology (the same as the mock catalogues
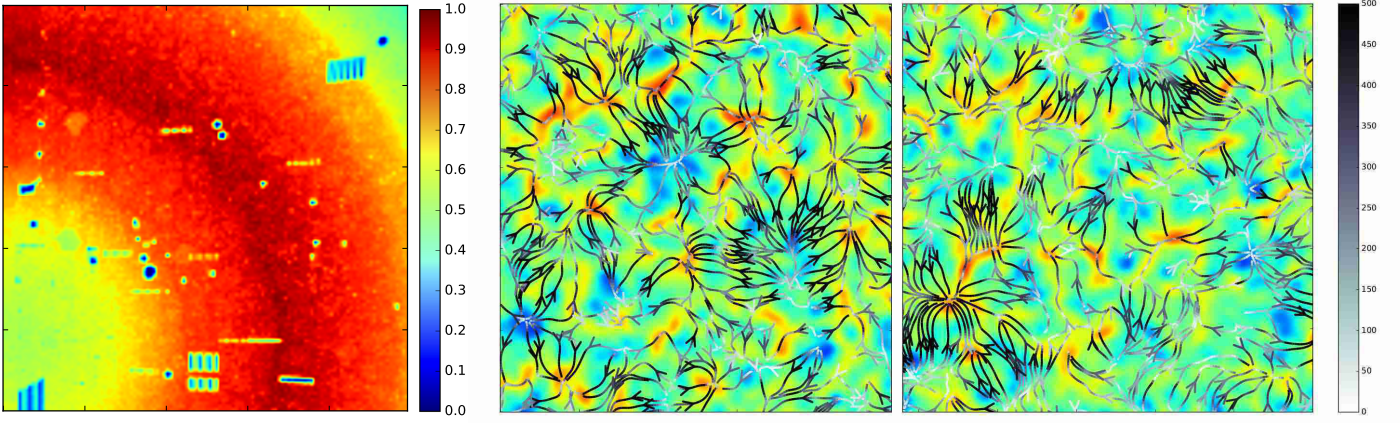
Figure 5. Slices of thickness $\sim 30\,h^{-1}$ Mpc in the $x - y$ plane of the 3D cubical mesh of side $1250\,h^{-1}$ Mpc and 200^3 cells, showing a zoom-in region of $900\,h^{-1}$ Mpc side for visual purposes. **Left panel:** the 3D completeness. Cosmic velocity fields with $6.25\,h^{-1}$ Mpc resolution with an additional Gaussian smoothing of the density and velocity field of $13\,h^{-1}$ Mpc smoothing radius based on (**middle panel:**) a light-cone mock catalogue in redshift space and on **right panel:** the BOSS DR12 data. The density of the stream lines corresponds to the field strength of the flows, whereas the color of the stream lines indicates its velocity at a particular position. The colour code for the density field is red for high and blue for low densities. A more quantitive comparison is shown in the figures below and in §4.3.
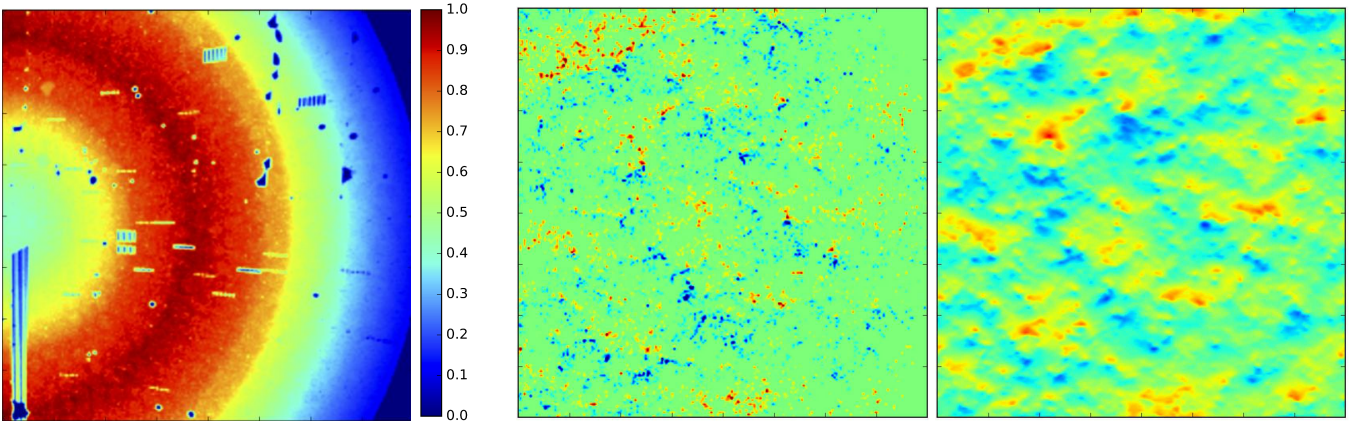
Figure 6. Based on a light-cone mock catalogue in redshift space with $6.25\,h^{-1}$ Mpc resolution and side $1250\,h^{-1}$ Mpc: slices in the $x - y$ plane of (**left panel:**) the 3D completeness, and the x-component of the velocity field for **middle panel:** the averaged mock galaxy velocities per cell, and **right panel:** one reconstructed velocity field sample with ARGO (compensating for completeness). The colour code for the density field is red for positive and blue for negative peculiar velocities. A more quantitive comparison is shown in Fig. 7.

described in Sec. 3.2), and transform angular coordinates (right ascension α and declination δ) and redshifts into co-moving Cartesian coordinates x, y, z

$$
\begin{aligned}
x &= r \cos \alpha \cos \delta \\
y &= r \sin \alpha \cos \delta \\
z &= r \sin \delta \,,
\end{aligned}
$$

with the comoving distance given by

$$
r = \frac{H_0}{c} \int_0^z \frac{dz'}{\sqrt{\Omega_{\rm M}(1+z')^3 + \Omega_\Lambda}} \,. \tag{41}
$$

With these transformations we can then grid the galaxies on a mesh and obtain the galaxy number count per cell $\boldsymbol{N}_{\rm G}$. In paticular, we consider in our analysis cubical volumes of L=1250 h^{-1} Mpc side length with 128^3 and 200^3 cells (and cubical volumes of L=3200 h^{-1} Mpc side length with 512^3

cells, see appendix A), and with the lower left corner of the box at

$$
\begin{aligned}
x_{\rm llc} &= -1500\,h^{-1}{\rm Mpc} \\
y_{\rm llc} &= -650\,h^{-1}{\rm Mpc} \\
z_{\rm llc} &= 0\,h^{-1}{\rm Mpc} \,.
\end{aligned}
$$

4.1.1 Completeness: angular mask and radial selection function

Furthermore, our data model requires the completeness in each cell $\boldsymbol{w}$ (see Eq. 17). The first ingredient in our three dimensional completeness is the angular mask of a position in the sky. The mask is provided as polygons with equal completeness (see left panel of Fig. 1, and the MANGLE software package, Hamilton & Tegmark 2004; Swanson et al. 2008). As a first step in the 3D completeness calculation we
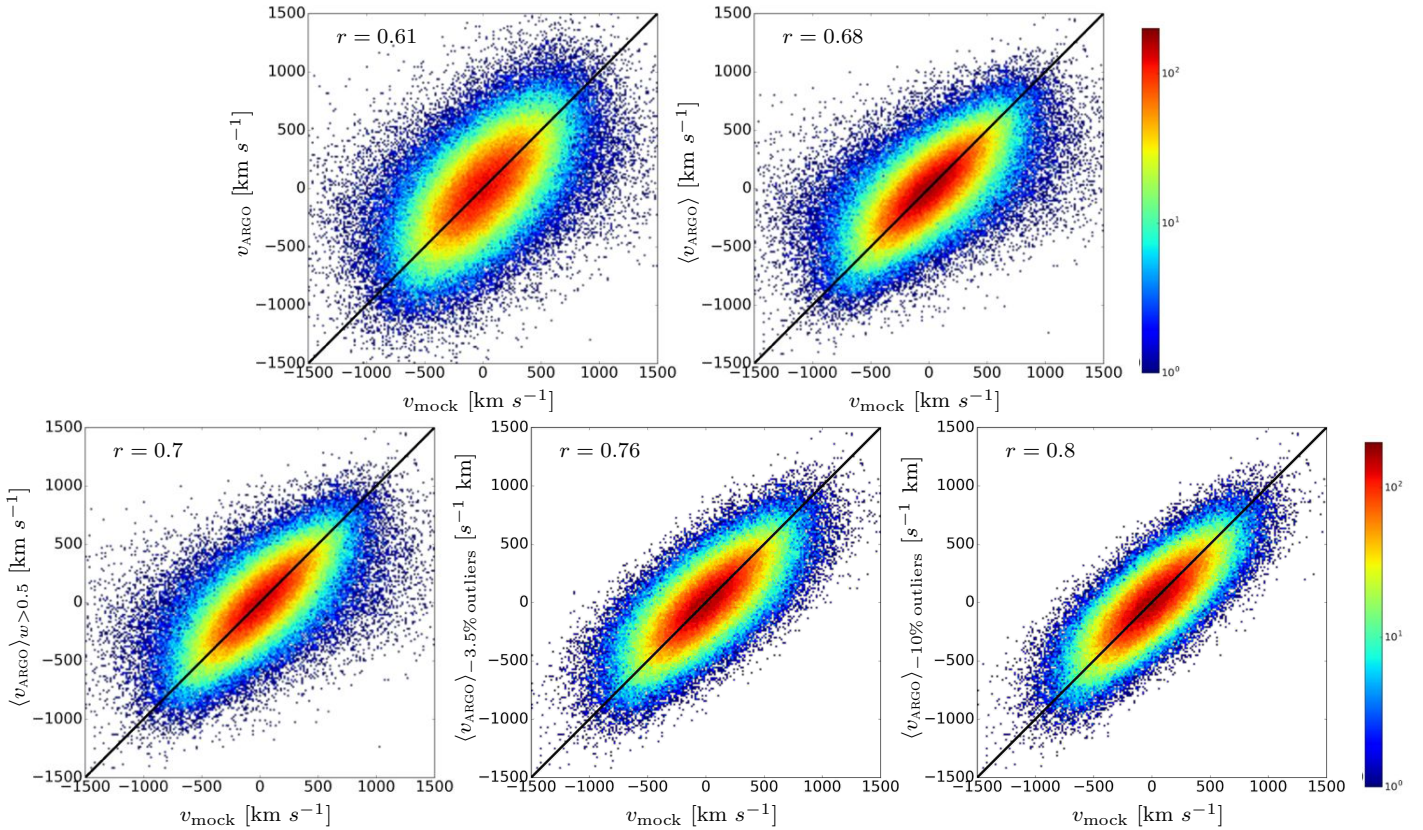
Figure 7. Velocity correlation taking one component of the velocity field for reconstructions with resolutions of $d_L = 6.25\,h^{-1}$ Mpc with additional Gaussian smoothing of $r_S = 2\,h^{-1}$ Mpc. **Upper left panel:** for one reconstructed sample, **upper right panel:** for the mean over 6000 reconstructed samples, **lower left panel:** same as upper right panel, but considering only galaxies with completeness $w > 0.5$ (for about 209000 galaxies, $\sim$82% of the whole CMASS sample in the considered volume), **lower middle panel:** same as upper right panel, but excluding galaxies for which the difference in the velocity reconstruction eceeds $|\boldsymbol{v}| = 700$ km s^{-1} (i.e., excluding about 3.5% of the sample), and **lower right panel:** same as upper right panel, but excluding galaxies for which the difference in the velocity reconstruction eceeds $|\boldsymbol{v}| = 500$ km s^{-1} (i.e., excluding about 10% of the sample).

project the angular mask to 3D by throwing large numbers of sight lines evaluating the sky mask with MANGLE. The result of such a projection is shown on the right panel of Fig. 1. Next we need to define the radial selection function from the number density distribution as a function of redshift

$$f(r) \propto \frac{1}{r^2} \frac{\Delta N_G}{\Delta r} \,. \tag{42}$$

normalised to one. In principle the radial selection function should be evualated in real space to avoid the so-called "Kaiser rocket" effect (Kaiser 1987; Nusser et al. 2014). This is only possible when the real space positions are reconstructed, as we do here. Obtaining the real space radial selection function can be expressed as an additional Gibbs-sampling step for iteration $j + 1$

$$f(r)^{j+1} \quad \curvearrowleft \quad \mathcal{P}_f\left(f(r)|\{r^j\}\right) \,, \tag{43}$$

for the set of recovered galaxy distances in the previous iteration $\{r^j\}$. Once we have the radial selection function we can multiply it with the 3D projected angular mask to get the 3D completeness. The radial selection functions as provided by the CMASS galaxy catalogue in redshift space and the reconstructed real space one are shown in Fig. 2. The agreement between both is very good, being compatible within 2-σ throughout almost the entire redshift range.

However, we see some tiny differences at distances where the selection function suffers strong gradients at the smallest distances, indicating that this approach could become important if such extreme cases happen more often.

4.2 Application to galaxy catalogues

In this section we present results from first testing the method on light-cone mocks resembling the BOSS CMASS survey geometry, radial selection function, and galaxy bias for which both the galaxy in real space and in redshift space are available; and second applying the same method to the BOSS DR12 data. We explore the scales between 6 and 10 h^{-1} Mpc. In particular, we consider grids with 128^3 and 200^3 cells with cubical volumes of L=1250 h^{-1} Mpc side. We need to obtain 12000 density and peculiar velocity samples using 8 cores for a mesh of 128^3 (200^3) cells about 200 (550) CPU hours with about <1 ($\sim$3) minute(s) per Gibbs-iteration given the survey geometry in this case study (less than 40% of the volume is covered with data). The memory requirements are 170 (870) Mb, respectively. An additional set of reconstructions considering volumes of 3200 h^{-1} Mpc and 512^3 cells has been done (see appendix A). We have chosen 10 redshift and completeness bins for the range in which
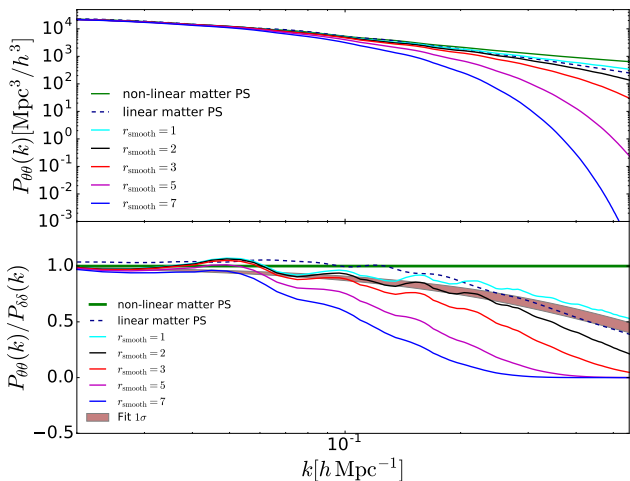
Figure 8. Upper panel: power spectrum of the scaled divergence of the peculiar velocity field for different smoothing scales for a typical realisation on a mesh of 200^3 with resolution $d_L = 6.25\,h^{-1}$ Mpc. **Lower panel:** ratio with respect to the non-linear power spectrum from Heitmann et al. (2010). The shaded region represents the theoretical fit for the velocity divergence bias $b_v = e^{-(k/a)^b}$ by Hahn et al. (2015) with the sigma region being computed based on the largest uncertainty found on the parameters a and b. The wiggles are due to the more pronounced baryon acoustic oscillations in the mean theoretical power spectrum than in the particular realisation used in this plot.

the CMASS data are defined in our study $0.43 < z < 0.7$ to sample the renormalisation of the lognormal fields (Eq. 12), and the normalisation of the number densities in the power law bias model (Eq. 18). A too fine resolution in redshift and completeness would introduce too much stochasticity in the derived μ and γ constants. We consider, however, 10 redshift bins to recover the peculiar velocity field at different redshifts (see Sec. 2.2). Here we do not take more redshift bins to save computational costs. In fact such a redshift spacing of 0.0225 is enough to model the cosmic evolution of CMASS galaxies (see Rodríguez-Torres et al. 2015; Kitaura et al. 2016a). The power spectrum correlation matrix shown in Fig. 3 demonstrates that after less than 200 iterations the chain converges to power spectra which are highly correlated. The correlation is less strong if one considers only the first 30 bins up to $k \sim 0.03\,h\,\mathrm{Mpc}^{-1}$, since cosmic variance due to empty and low completeness regions in the volume dominates those scales. Nevertheless, even on those scales we expect to have high correlations between power spectra of different iterations after convergence due to the constrained phases by the data. It is therefore safe to disregard the first 1000 iterations of the chains until the power spectra have converged and use a total of 6000 iterations for our analysis for each setup (meshes of 128^3 and 200^3 for mocks and observations). We further demonstrate that we succeed in sampling from the posterior distribution function estimated through the Gelman & Rubin (1992) test as shown on the right panel of Fig. 3 (for details, see appendix in Ata et al. 2015). The linear real space bias b_L, one typical sample of the lognormal mean field $\mu = \langle \log(1 + \delta) \rangle$, and of the galaxy density normalisation γ, are shown in Fig. 4. We find that it is crucial to sample the bias and

the mean fields on at least 5 bins to get accurate density reconstructions free of radial selection biases. However, the reconstructions are robust against different redshift bins in γ. We find that the theoretical prediction for the mean field $\mu = -\sigma^2/2 \simeq -0.760$ for resolutions of $6.25\,h^{-1}$ Mpc is compatible within 4% with our numerical sampling result. The technique presented in this work permits us to get peculiar velocity fields which are compensated for the survey geometry and selection functions. This can be qualitatively appreciated in Figs. 5 and 6. On a quantitative level, we find that the velocities are highly correlated and approximately unbiased with the true velocities (see Fig 7) for the case in which the density fields was smoothed with a Gaussian kernel with radius of $r_S = 2\,h^{-1}$ Mpc for a resolution of $d_L = 6.25\,h^{-1}$ Mpc. We note that the maximum a posteriori (MAP) solution, such as Wiener-filtering, will yield biased results, although for Wiener-filtering the variance can be separately added to the MAP solution and such a bias is known (Zaroubi et al. 1995). The statistical correlation coefficient we find is about 0.7 including about 10% of satellite galaxies with virial motions, which is what one finds for CMASS galaxies. We have checked this result testing boundary effects and cosmic variance by considering the full volume covered by the CMASS sample (see appendix A). We find very similar results to the sub-volume reconstructions, which at most decrease the statistical correlation coefficient to about 0.69. This correlation can be considerably improved by excluding these satellite galaxies from the analysis. As a proxy we consider two cases. One excluding galaxies for which the velocity difference between true and reconstructed exceeds 500 and 700 km s^{-1}. The first one removes ∼10% of the galaxies, and the second one ∼3.5%. Since not all satellite galaxies will be outliers the answer will be probably closer to the latter case, rasing the statistical correlation coefficient to about $r = 0.75$, which is a priori a considerable improvement with respect to previous methods (see, e.g., Schaan et al. 2015; Planck Collaboration et al. 2016, though a proper comparison between methods remains to be done based on the same mocks). Although we are using only linear theory here, our method includes a couple of ingredients which can explain this improvement, such as being a self-consistent (iterative) method, yielding linearised density fields, for which the pixel window has been exactly solved (the counts in cells, i.e., the nearest grid point, are treated through the full Poisson likelihood), and nonlinear bias has been taken into account. The smoothing scale could be considered another parameter of our model. However, it can be derived from the velocity divergence power spectrum $P_{\theta\theta}$ with $\theta \equiv -\frac{1}{fHa}\nabla \cdot \boldsymbol{v}$ prior to running any Markov chain, as it has been done here. In particular, one expects $P_{\theta\theta}$ to converge towards the linear power spectrum in the transition to the nonlinear regime at about $k \sim 0.15 - 0.2\,h\,\mathrm{Mpc}^{-1}$ (Jennings 2012; Hahn et al. 2015). This is expected as the velocity divergence is closer to the Gaussian field than the gravitationally evolved density field (see, e.g., Kitaura et al. 2012b). In fact while the density is enhanced in the potential wells, virialisation prevents galaxies from getting larger and larger velocities. As a consequence, the power spectrum of the velocity divergence is close to the linear density field in the quasi-linear regime, eventually being even more suppressed at high k values. Fig. 8 shows that such an agreement down to scales of $k \sim 0.2\,h\,\mathrm{Mpc}^{-1}$ is indeed achieved
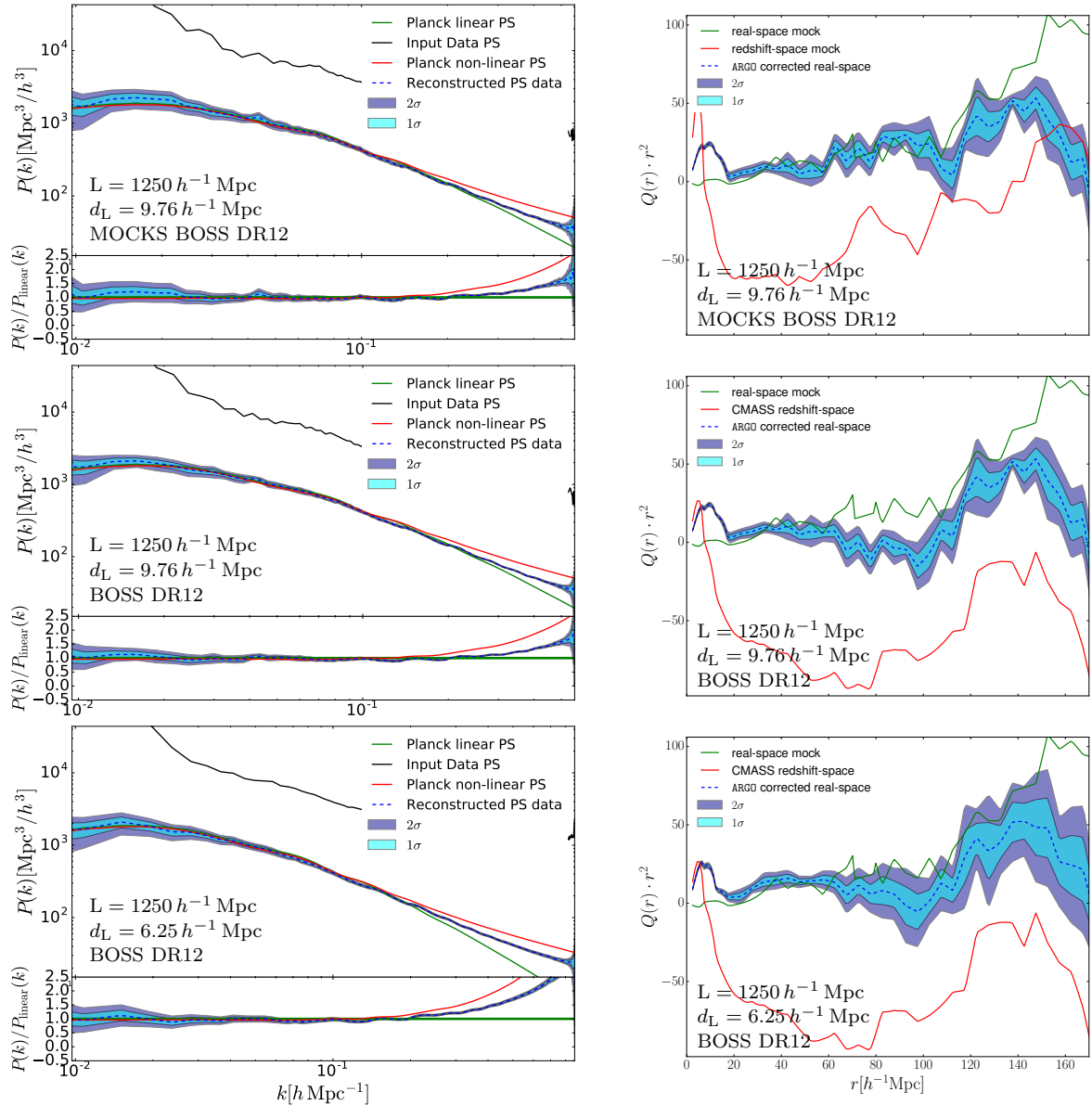
Figure 9. Left panels: power spectra of the reconstructed density fields $\delta(z_{\mathrm{ref}})$ on a mesh and **right panels:** quadrupoles of the galaxy distribution $\{s^{\mathrm{obs}}\}$ and $\{r\}$ based on **upper panels**: a light-cone mock (including survey geometry) with $d_{\mathrm{L}} = 9.76\,h^{-1}$ Mpc, **middle panels**: the BOSS DR12 data with $d_{\mathrm{L}} = 9.76\,h^{-1}$ Mpc, and **lower panels**: the BOSS DR12 on with $d_{\mathrm{L}} = 6.25\,h^{-1}$ Mpc. Power spectra show the mean (dashed blue line) over 6000 samples with 1 and 2 σ contours (light and dark blue shaded areas, respectively), as compared to the raw galaxy power spectrum (black solid line), the nonlinear (red solid line), and the linear power spectrum (green solid line) assuming the fiducial cosmology. Quadrupole correlation functions show the mean (dashed blue line) over 6000 samples (10 spaced samples in intervals from 500 iterations covering 4000 Gibbs-iterations for quadrupoles to reduce computations) with 1 and 2 σ contours (light and dark blue shaded areas, respectively), as compared to the raw galaxy power spectrum (black solid line), and the corresponding computations for the catalogues in real (green line for mocks only) and redshift space (red line).

for smoothing scales of about $r_{\mathrm{S}} = 2\,h^{-1}$ Mpc. In fact for a smoothing scale r_{S} between 1 and 2 h^{-1} Mpc one can potentially obtain unbiased results beyond $k = 0.5\,h$ Mpc^{-1}. While our chains with 128^3 were run with velocities derived from density fields smoothed with $r_{\mathrm{S}} = 7\,h^{-1}$ Mpc, our reconstructions with 200^3 were run using $r_{\mathrm{S}} = 2\,h^{-1}$ Mpc. This variety of smoothing scales serves us to test the robustness of the velocity reconstructions depending on this parameter. In fact we manage to recover the monopoles in real space down to scales of about $k \sim 0.2\,h$ Mpc^{-1} (see left

panels in Fig. 9 and 10, for the lognormal-Poisson). We have checked that the theoretical prediction from renormalised perturbation theory for the bias correction parameter f_{b} can be sampled as a free parameter yielding compatible results, $f_{\mathrm{b}} = 0.70 \pm 0.05$ vs $f_{\mathrm{b}} = 0.66 \pm 0.1$ from theory when considering the first 30 bins in the power spectrum, i.e., $k \lesssim 0.03\,h$ Mpc^{-1}. Given the volume we consider in this work of $(1250\,h^{-1}$ Mpc$)^3$ we expect cosmic variance to cause deviations from zero in the quadrupoles. Therefore, we show the quadrupole of the real space mock galaxy catalogue as a
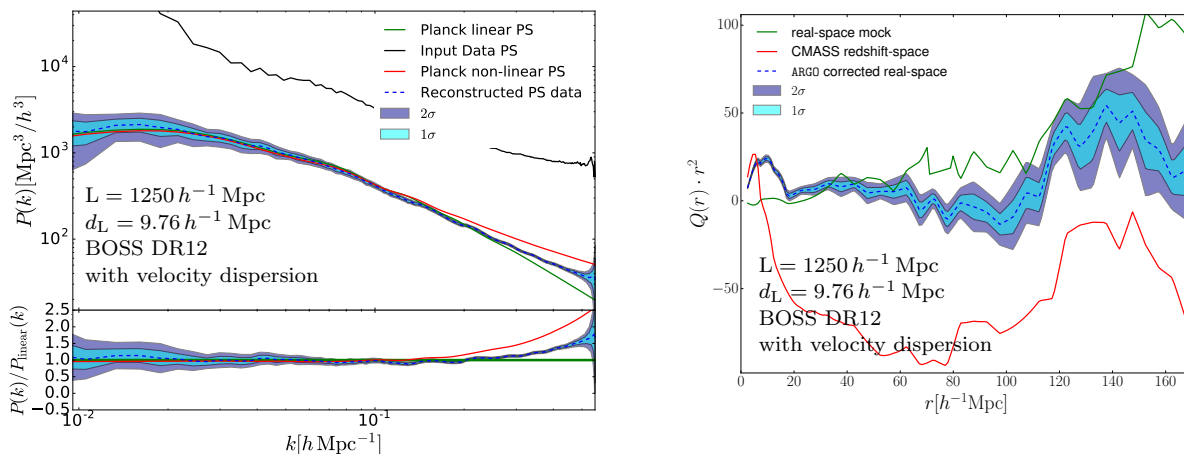
Figure 10. Same as Fig. 9 but including velocity dispersion.

reference. The upper right panel in Fig. 9 demonstrates that we cover the real space quadrupole down to scales of about $r \sim 20\,h^{-1}$ Mpc. Deviations on large scales ($\gtrsim 120\,h^{-1}$ Mpc) between the recovered and the true quadrupoles are due to the large empty volume which pushes the solution to be closer to zero than in the actual mock catalogue. In fact, we showed in a previous paper that one can recover with this method the quadrupole features of the particular realisation when considering complete volumes (Kitaura et al. 2016b). The results are consistent when comparing lower to higher resolution reconstructions (middle to lower right panels). However, we see that the uncertainty (shaded regions) in the quadrupole increases in the higher resolution case. This is expected as the coarser grid smooths the peculiar velocities and tends to underestimate them. In addition, we have run a reconstruction chain including velocity dispersion, showing that this will also enhance the error bars in the quadrupole, however yielding the same qualitative results as without that term (see lower panels in Fig. 10). We observe a slightly enhanced uncertainty in the monopole and quadrupole on large scales. A proper treatment of the velocity dispersion requires, however, at least a density dependent dispersion term, or even looking at the tidal field Eigenvalues (see Kitaura et al. 2016b). This is however, computationally more expensive and requires a number of additional parameters. We thus leave such an effort for later work. The accuracy of the quadrupole reconstruction presented in this paper seems to be superior than in some of the standard BAO reconstruction techniques (Vargas-Magaña et al. 2015; Burden et al. 2015), see in particular, right panel in Fig. 8 in Kitaura et al. (2016a) showing the quadrupole after BAO reconstruction for a set of mock Multidark-PATCHY BOSS DR12 CMASS catalogues very similar to the ones used here. We note that while the monopoles of the dark matter field are trivially computed from the reconstructed samples on complete meshes, the computation of the quadrupoles of the galaxies needs more computational efforts to account for survey geometry and radial selection functions (see, e.g., Anderson et al. 2014a). Fig. 11 shows slices in the $x - z$ plane of the galaxy number counts, the completeness, and the reconstructed density fields. One can clearly recognise prominent features in the data in the reconstructed density

fields. It is remarkable however, how these features appear balanced without selection function effects, in such reconstructions. Only when one computes the mean over many realisations, one can see that larger significance in the reconstructions correlates with higher completeness values. The vanishing structures in unobserved regions further demonstrates the success in sampling from the posterior distribution function. Fig. 12 shows that the lognormal fields are indeed reasonably Gaussian distributed in terms of the univariate probablity distribution function. In fact the absolute skewness is reduced from about 6.4 to less than 0.03 with means being always smaller than $|\langle \delta_{\rm L} \rangle| < 0.13$ for different completeness regions. As we will analyse below the 3pt statistics does, however, not correspond to a Gaussian field.

4.3 The cosmic web from lognormal-Poisson reconstructions

So far we have been reconstructing the linear component of the density field in Eulerian space at a reference redshift within the lognormal approximation. We can, however, get an estimate of the nonlinear cosmic web by performing structure formation within a comoving framework, i.e., without including the displacement of structures, as our reconstructed linear density fields already reside at the final Eulerian coordinates. One can use cosmological perturbation theory to make such a mapping (see Kitaura & Angulo 2012). We will rely here on the classical Zel'dovich (1970) framework. By demanding mass conservation from Lagrangian to Eulerian space $\rho(\boldsymbol{q})\mathrm{d}\boldsymbol{q} = \rho(\boldsymbol{r})\mathrm{d}\boldsymbol{r}$, we get an equation for the cosmic evolved density field within comoving coordinates: $1 + \delta^{\rm PT}(\boldsymbol{q}) = \mathbf{J}^{-1}$ (with the supercript standing for perturbation theory), where $\mathbf{J}$ is the Jacobian matrix often called the tensor of deformation: $\mathcal{D}_{ij} \equiv \delta^{\rm K}_{ij} + \Psi_{i,j}(\boldsymbol{q}, z)$. By doing the proper diagonalisation one finds that the comoving evolved density field can be written as:

$$\delta^{\rm PT}(\boldsymbol{q}, z) = \frac{1}{(1 - D(z)\lambda_1(\boldsymbol{q}))(1 - D(z)\lambda_2(\boldsymbol{q}))(1 - D(z)\lambda_3(\boldsymbol{q}))} - 1\,,$$
(44)

where λ_i are the Eigenvalues of the deformation tensor with $\lambda_1 \geq \lambda_2 \geq \lambda_3$. This framework is helpful to gain insight
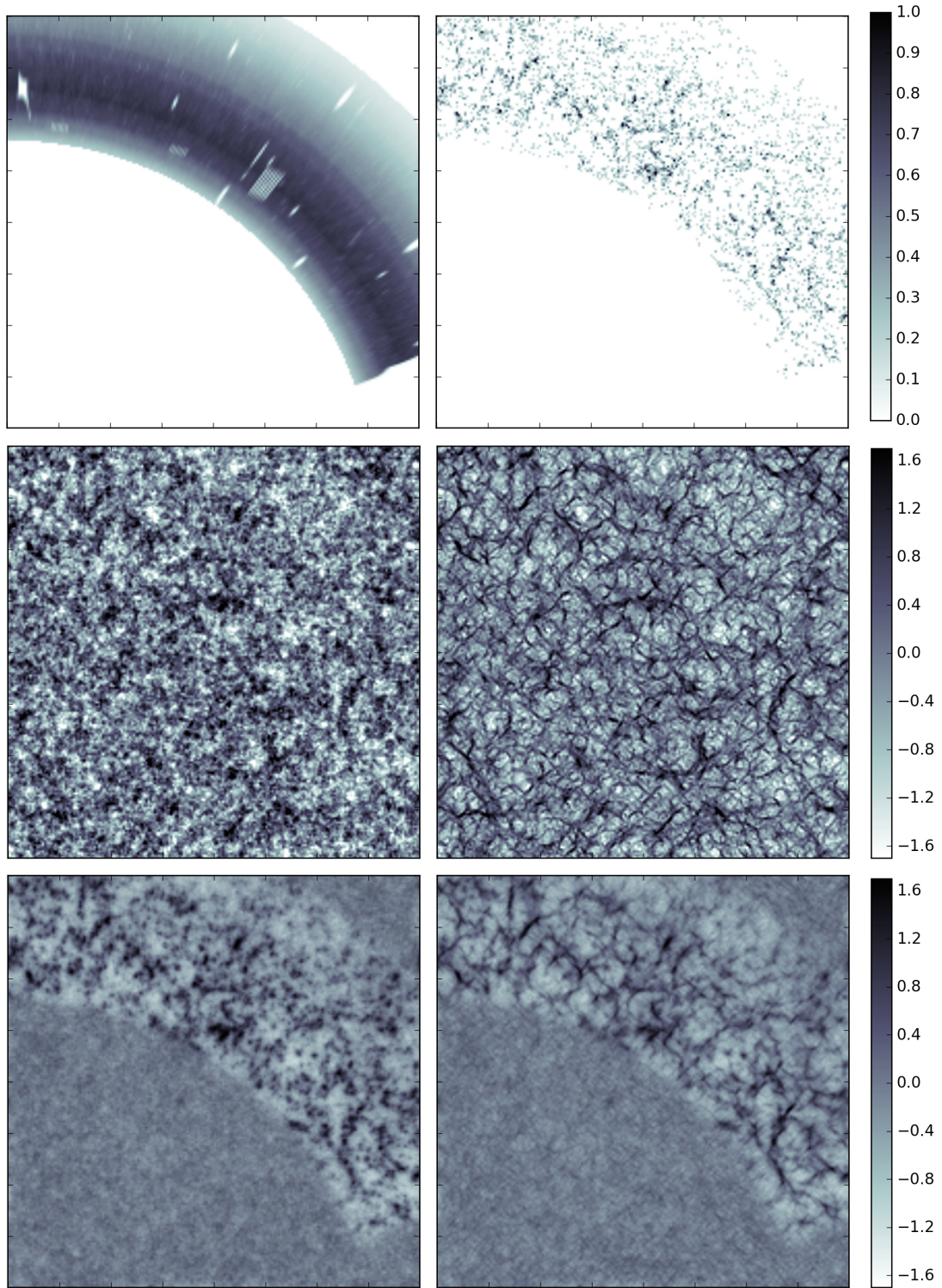
Figure 11. Based on a light-cone mock catalogue in redshift space with $6.25\,h^{-1}$ Mpc resolution: slices of thickness $\sim 30\,h^{-1}$ Mpc in the $x-z$ plane of the 3D cubical mesh of side $1250\,h^{-1}$ Mpc and 200^3 cells for the following quantities: **upper left panel:** the 3D completeness or window function multiplied with a factor of 0.8 for visualisation purposes, **upper right panel:** the number counts per cell in real space, **middle left panel:** one reconstructed linear logarithmic density sample of the lognormal-Poisson field, **middle right panel:** same as middle left panel for the Zeldovich transformed density, **lower left panel:** the mean over the linear logarithmic density sample over 6000 reconstructions of the lognormal-Poisson field, and **lower right panel:** same as lower left panel for the Zeldovich transformed density.
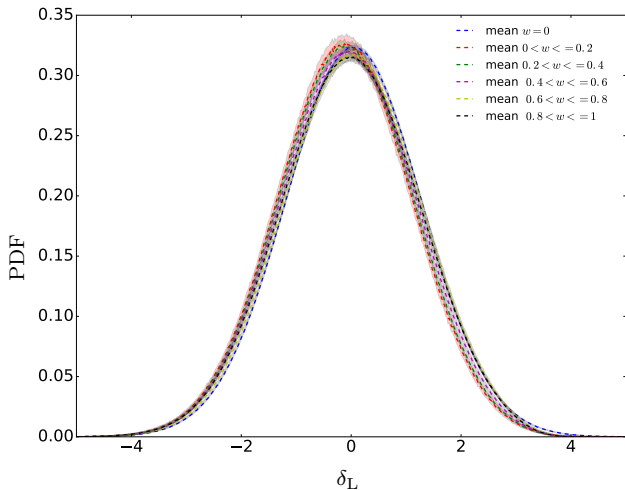
Figure 12. PDF of the matter statistics for different completeness values from 6000 reconstructions on a mesh of 200^3 and resolution $d_{\rm L} = 6.25\,h^{-1}$ Mpc for the linear component reconstructed with the lognormal-Poisson model. The corresponding skewness range between -10^{-4} and -0.09 with means being always smaller than $|\langle \delta_{\rm L} \rangle| < 0.13$. The skewness is thus reduced by two orders of magnitude, as compared to a skewness of ~ 7 corresponding to the galaxy overdensity on a mesh with a cell resolution of $10\,h^{-1}$ Mpc.

over the formation of the cosmic web (see Hahn et al. 2007). In fact we could use the reconstructed velocity field to compute the shear tensor and study the cosmic web (Bond et al. 1996). We will however, focus on the largest Eigenvalue denoting the direction of first collapse to form the filamentary cosmic web. We can Taylor expand the previous equation within the Eulerian framework yielding

$$\delta^{\rm PT}(\boldsymbol{r}, z) \simeq D(z)\lambda_1(\boldsymbol{r}) + \lambda^+(\boldsymbol{r}, z)\,, \qquad (45)$$

with λ^+ being the higher order contributions including the rest of Eigenvalues, which can be approximated by $\lambda^+(\boldsymbol{r}, z) \simeq -\langle D(z)\lambda_1(\boldsymbol{r})\rangle$. This expression avoids the problem of formation of caustics, as present in Eq. 44. We have tested other expansions including the rest of Eigenvalues, however, with less success in describing the nonlinear cosmic web. The operation of retaining the information of the largest Eigenvalue can also be interpreted, as filtering out the noisy part of the Gaussian field. This technique could potentially be useful to effectively enhance the cosmic web of a low resolution simulation for mock catalogue production. We leave a more thorough investigation of other possible comoving structure formation descriptions for later work. Since this theory is based on the Gaussian density field, we will compute the Eigenvalues based on the linear component of the density field $\delta_{\rm L}$. In particular, we will compute them from the gravitational potential $\phi_{\rm L} \equiv \nabla^{-2}\delta_{\rm L}$, solving the Poisson equation with the inverse Laplacian operator in Fourier space, to obtain the corresponding tidal field tensor. By applying Eq. 45 we thus get the linear component of the gravitationally evolved density field in Eulerian space, which we will denote as $\delta_{\rm L}^{\rm PT}(\boldsymbol{r})$. We now can compute the nonlinear component by doing the transformation $\delta^{\rm PT}(\boldsymbol{r}) = \exp(\delta_{\rm L}^{\rm PT}(\boldsymbol{r}) + \mu(\delta_{\rm L}^{\rm PT}(\boldsymbol{r})) - 1$, having the physical meaninful property of yielding positive definite density fields. To ensure that this field shares the same

power spectrum, as the lognormal reconstructed density field $\delta(\boldsymbol{r}) = \exp(\delta_{\rm L}(\boldsymbol{r}) + \mu(\delta_{\rm L}(\boldsymbol{r}))) - 1$, we apply in Fourier space

$$\hat{\delta}_{\rm L}^{\rm PT,f}(\boldsymbol{k}) = \sqrt{P^{\rm trans}(k)} \frac{\hat{\delta}_{\rm L}^{\rm PT}(\boldsymbol{k})}{\sqrt{\langle |\hat{\delta}_{\rm L}^{\rm PT}(\boldsymbol{k})|^2 \rangle_{\Delta k}}}\,, \qquad (46)$$

where the nonlinear transformed power spectrum $P^{\rm trans}(k)$ is found iteratively. The ratio between the target power spectrum and the one obtained at a given iteration is multiplied to $P^{\rm trans}(k)$ from the previous iteration until the nonlinear power spectra averaged in Δk-shells coincide $\langle |\hat{\delta}^{\rm PT,f}(\boldsymbol{k})|^2 \rangle_{\Delta k} \simeq \langle |\hat{\delta}(\boldsymbol{k})|^2 \rangle_{\Delta k}$ (i.e., the power spectrum from the nonlinear transformed lognormal density field), in a given k-range within a given accuracy. As a starting guess of $P^{\rm trans}(k)$ we take $\langle |\hat{\delta}_{\rm L}(\boldsymbol{k})|^2 \rangle_{\Delta k}$ (i.e., the power spectrum from the linear lognormal density field). In practice, less than 15 iterations are necessary to be accurate within better than 1% up to at least 70% of the Nyquist frequency using about 100 Δk-bins for meshes of 200^3 cells on cubical volumes of $1250\,h^{-1}$ Mpc side, requiring less than $100s$ on 8 cores. This operation is justified, as we are dealing with the Gaussian component of the density field, permitting us to define a pseudo white noise $\hat{\delta}_{\rm L}^{\rm PT}(\boldsymbol{k})/\sqrt{\langle |\hat{\delta}_{\rm L}^{\rm PT}(\boldsymbol{k})|^2 \rangle_{\Delta k}}$, which allows modifications of the two point statistics. In fact, the PDF of $\delta_{\rm L}^{\rm PT}$ is very Gaussian. This calculation is parameter free, and does not require any further input than the lognormal field (and the window function to compute the completeness dependent renormalised mean fields). Effectively, these transformations retain the two-point statistics, while improving the three-point statistics of the lognormal field, hereby extracting the cosmic web structure of the density field, which is diluted in the lognormal reconstructions (for a similar concept see Leclercq et al. 2013). We note, that the distribution of peaks even prior to the nonlinear tidal field transformation do not correspond to a random lognormal realisation, as they are based on the galaxy distribution within the posterior sampling analysis, which already suffered displacements due to the action of gravity. The results of this study are shown in Figs. 11, 13, 14, and 15. One can see how the closely Gaussian logarithmic-density field (lower left panels in Figs. 11 and 13) is transformed into a density field depicting the cosmic web (lower right panels in Figs. 11 and 13), which is in good agreement with the distribution of galaxies (see Fig. 14). The ensemble average plots shown in the lower right panels of Figs. 11 and 13 demonstrate the robustness of the reconstructed filamentary network. The variance plots confirm as expected that the uncertainty on the density is larger in the voids and in the unobserved regions (see lower panels in Fig. 15). In fact, the variance depicts the negative of the filamentary network. These density maps can be used for environmental studies (see, e.g., Nuza et al. 2014). They could be used as a reference for future applications including reconstructions of the initial conditions (see e.g. Kitaura et al. 2012c; Jasche & Wandelt 2013; Kitaura 2013; Wang et al. 2013; Heß et al. 2013; Wang et al. 2014).
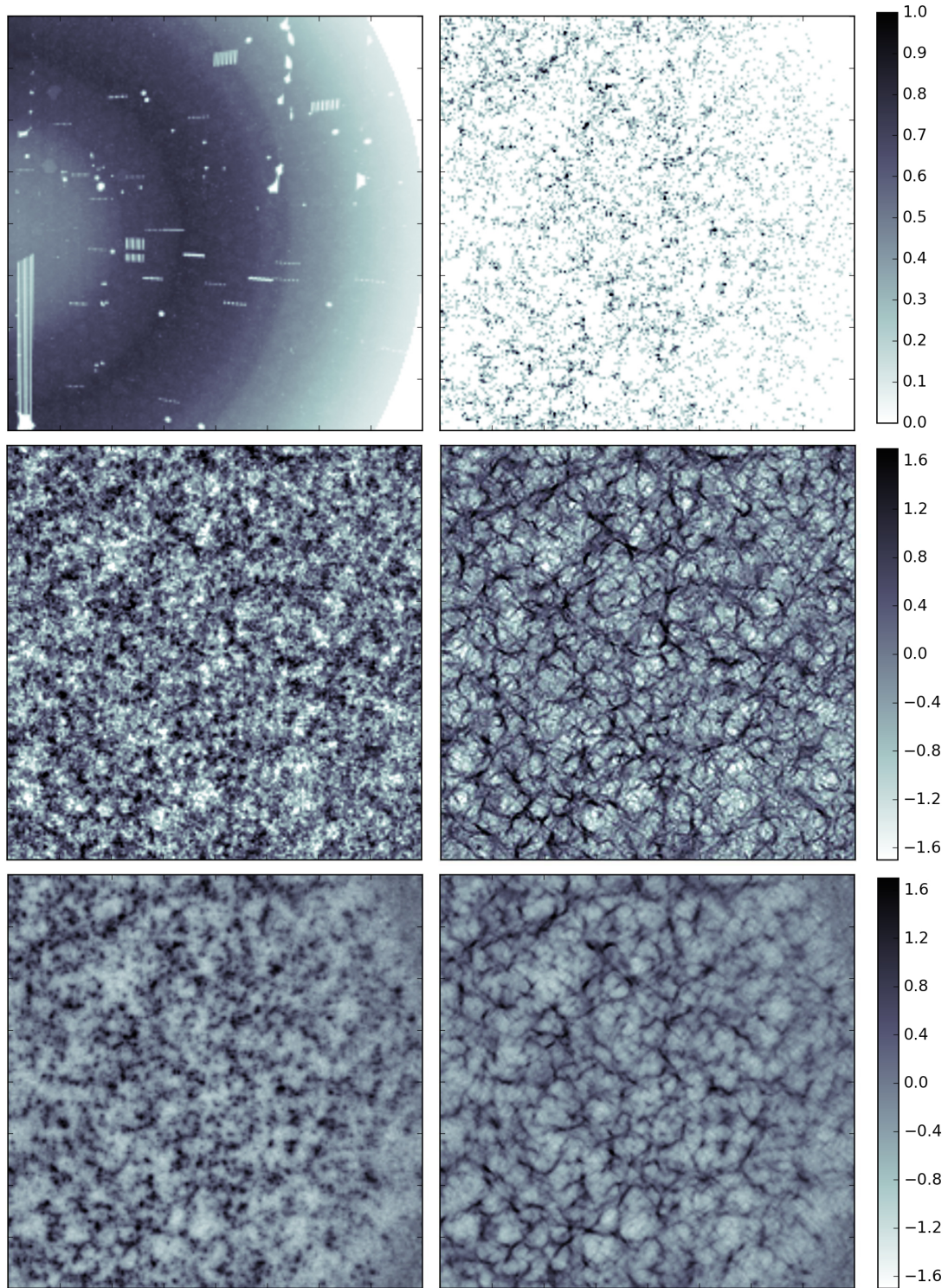
Figure 13. Same as Fig. 11, but for the $x - y$ plane.

5   SUMMARY AND CONCLUSIONS

In this work, we have presented a Bayesian phase-space (density and velocity) reconstruction of the cosmic large-scale matter density and velocity field from the SDSS-III Baryon Oscillations Spectroscopic Survey Data Release 12 (BOSS DR12) CMASS galaxy clustering catalogue. We have demonstrated that very simple models can yield accurate results on scales larger than $k \sim 0.2 \, h \, \mathrm{Mpc}^{-1}$.

In particular we have used a set of simple assumptions. Let us list them here

- the statistical distribution of galaxies is described by the lognormal-Poisson model,
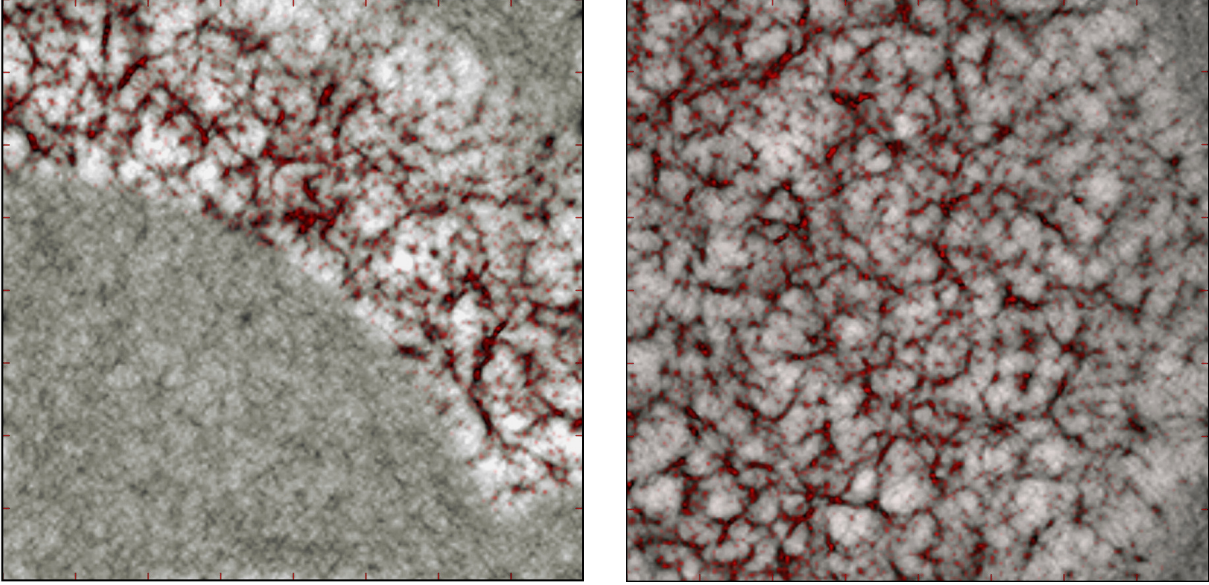
Figure 14. Slices of the ensemble averaged Zeldovich transformed density field shown in Figs. 11 (left) and 13 (right) with another colour bar for visualisation purposes and the corresponding real space galaxy number count per cell overplotted in red.
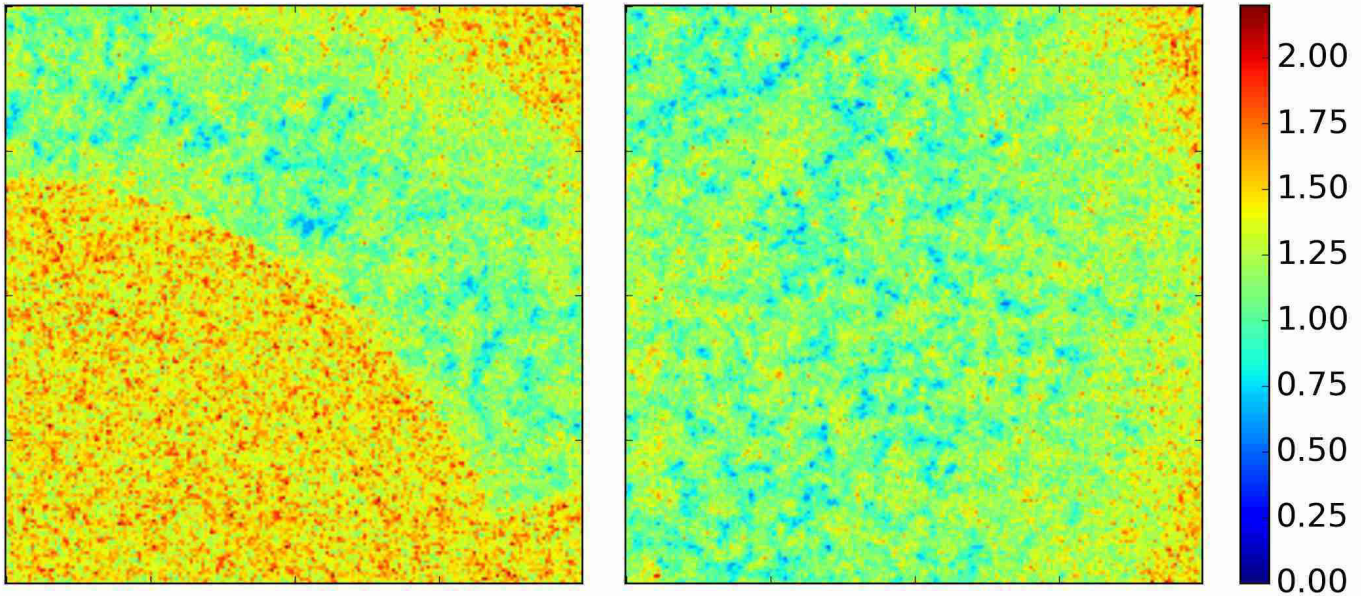
Figure 15. Slices of the variance corresponding to the **(left panel:)** $x - z$ plane, and **(right panel:)** $x - y$ plane shown in Figs. 11 and 13, respectively.

- linear theory relates the peculiar velocity field to the density field,
- the volume is a fair sample, i.e. ensemble averages are equal to volume averages,
- cosmic evolution is modelled within linear theory with redshift dependent growth factors, growth rates, and bias,
- a power law bias, based on the linear bias multiplied by a correction factor, which can be derived from renormalised perturbation theory, relates the galaxy expected number counts to the underlying density field.

This has permitted us to reduce the number of parameters and derive them consistently from the data, with a given smoothing scale and a particular ΛCDM cosmological parameter set.

We have included a number of novel aspects in the ARGO code extending it to account for cosmic evolution in the linear regime. In particular, the Gibbs-scheme samples

- the density fields with a lognormal-Poisson model,
- the mean fields of the lognormal renormalised priors for different completeness values,
- the number density normalisation at different redshift bins,
- the real space positions of galaxies from the reconstructed peculiar velocity fields,

• and the real space radial selection function from the reconstructed real space positions of galaxies (accounting for the "Kaiser-rocket" effect).

Our results show that we can get unbiased dark matter power spectra up to $k \sim 0.2\,h\,\mathrm{Mpc}^{-1}$, and unbiased isotropic quadrupoles down to scales of about $20\,h^{-1}$ Mpc, being far superior to redshift space distortion corrections based on traditional BAO reconstruction techniques which start to deviate at scales below $60\,h^{-1}$ Mpc.

As a test case study we also analyse deviations of Poissonity in the likelihood, showing that the power in the monopole and the scatter in the quadrupoles is increased towards small scales.

The agreement between the reconstructions with mocks and BOSS data is remarkable. In fact, the identical algorithm with the same set-up and parameters were used for both mocks and observations. This confirms that the cosmological parameters used in this study are already close to the true ones, the systematics are well under control, and gives further support to ΛCDM at least on scales of about $0.01 \lesssim k \lesssim 0.2\,h\,\mathrm{Mpc}^{-1}$.

We also found that the reconstructed velocities have a statistical correlation coefficient compared to the true velocities of each individual lightcone mock galaxy of $r \sim 0.7$ including about 10% of satellite galaxies with virial motions. The power spectra of the velocity divergence agree well with theoretical predictions up to $k \sim 0.2\,h\,\mathrm{Mpc}^{-1}$. This is far superior to the results obtained from simple linear reconstructions of the peculiar velocities directly applied on the smoothed galaxy field for which statistical correlation coefficients of the order of 0.5 are obtained (Planck Collaboration et al. 2016, though this work used the Sloan main sample at lower redshifts being further in the nonlinear regime, making a direct comparison difficult). Improved results can be obtained with Wiener-filter based techniques, which need to correct for the bias in a post-processing way (Schaan et al. 2015). It would be interesting to compare the different methods, in particular considering that the ensemble average is not equal to the maximum of the posterior for non-Gaussian PDFs, as we consider here. Although it may seem surprising to get such accurate results from simply assuming linear theory to derive the peculiar motions, we expect that linearised density fields as the ones obtained from lognormal-Poisson reconstructions (even if one takes the nonlinear transformed one), yield improved velocity fields (see Falck et al. 2012; Kitaura & Angulo 2012). Also, while linear theory tends to overestimate the peculiar velocity field, the chosen grid resolution with the additional smoothing compensates for this yielding unbiased reconstructed peculiar motions. We have seen that for a given resolution the additonal Gaussian smoothing radius (and the cell resolution) can be derived from the velocity divergence power spectrum to match the linear power spectrum in the quasi-linear regime ($0.1 \lesssim k \lesssim 0.5\,h\,\mathrm{Mpc}^{-1}$). We demonstrated that the reconstructed linear component reduces the skewness by two orders of magnitude with respect to the density directly derived from smoothing the galaxy field on the same scale.

We have furthermore demonstrated how to compute the Zeldovich density field from the lognormal reconstructed density fields based on the tidal field tensor in a parame-ter free way. The recovered filamentary network remarkably connects the discrete distribution of galaxies. The real space density fields obtained in this work could be used to recover the initial conditions with techniques which rely on knowing the dark matter field at the final conditions (see e.g. Wang et al. 2013, 2014).

We aim to improve the Bayesian galaxy distance estimates going to smaller scales, by using non-Poisson likelihoods and including a correction of the virialised motions (Ata et al. 2015; Kitaura et al. 2016b). One could also explore other priors based on perturbation theory (e.g. Kitaura & Heß 2013; Heß et al. 2013).

Despite of the potential improvements to this work, the reconstructed density and peculiar velocity fields obtained here can already be used for a number of studies, such as BAO reconstructions, kinematic Sunyaev-Zeldovich (kSZ), integrated Sachs-Wolfe (ISW) measurements, or environmental studies.

REFERENCES

Ahn K., Iliev I. T., Shapiro P. R., Srisawat C., 2015, MNRAS, 450, 1486
Alam S., et al., 2015a, ApJS, 219, 12
Alam S., Ho S., Vargas-Magaña M., Schneider D. P., 2015b, MNRAS, 453, 1754
Anderson et al. L., 2014a, MNRAS, 441, 24
Anderson L., et al., 2014b, MNRAS, 441, 24
Ata M., Kitaura F.-S., Müller V., 2015, MNRAS, 446, 4250
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Bel J., et al., 2014, A&A, 563, A37
Benitez N., et al., 2014, preprint, (arXiv:1403.5237)
Berlind A. A., Narayanan V. K., Weinberg D. H., 2001, ApJ, 549, 688
Bernardeau F., 1992, ApJ, 390, L61
Bernardeau F., Chodorowski M. J., Łokas E. L., Stompor R., Kudlicki A., 1999, MNRAS, 309, 543
Beutler F., et al., 2014, MNRAS, 443, 1065
Bilicki M., Chodorowski M. J., 2008, MNRAS, 391, 1796
Blake C., et al., 2011, MNRAS, 415, 2876
Blake C., et al., 2013, MNRAS, 436, 3089
Bolton A. S., et al., 2012, AJ, 144, 144
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
Branchini E., Eldar A., Nusser A., 2002, MNRAS, 335, 53
Branchini E., Davis M., Nusser A., 2012, MNRAS, 424, 472
Burden A., Percival W. J., Howlett C., 2015, MNRAS, 453, 456
Carron J., Szapudi I., 2014, MNRAS, 439, L11
Cen R., Ostriker J. P., 1993, ApJ, 417, 415
Chodorowski M. J., Lokas E. L., Pollo A., Nusser A., 1998, MNRAS, 300, 1027
Chuang C.-H., Wang Y., 2013a, MNRAS, 431, 2634
Chuang C.-H., Wang Y., 2013b, MNRAS, 435, 255
Chuang C.-H., et al., 2013a, preprint, (arXiv:1312.4889)
Chuang C.-H., et al., 2013b, MNRAS, 433, 3559
Chuang C.-H., et al., 2015, MNRAS, 452, 686
Cole S., Fisher K. B., Weinberg D. H., 1995, MNRAS, 275, 515
Coles P., Jones B., 1991, MNRAS, 248, 1
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Courtois H. M., Hoffman Y., Tully R. B., Gottlöber S., 2012, ApJ, 744, 43
Croft R. A. C., Gaztanaga E., 1997, MNRAS, 285, 793
Davis M., Nusser A., Willick J. A., 1996, ApJ, 473, 22
Dawson K. S., et al., 2013, AJ, 145, 10
Duane S., Kennedy A. D., Pendleton B. J., Roweth D., 1987, Physics Letters B, 195, 216
Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007, ApJ, 664, 675
Eisenstein D. J., et al., 2011, AJ, 142, 72
Erdogdu P., et al., 2004, MNRAS, 352, 939
Falck B. L., Neyrinck M. C., Aragon-Calvo M. A., Lavaux G., Szalay A. S., 2012, ApJ, 745, 17
Fisher K. B., Lahav O., Hoffman Y., Lynden-Bell D., Zaroubi S., 1995, MNRAS, 272, 885
Frieman J., Dark Energy Survey Collaboration 2013, in American Astronomical Society Meeting Abstracts. p. 335.01
Fry J. N., Gaztanaga E., 1993, ApJ, 413, 447

Gelman A., Rubin D., 1992, Statistical Science, 7, 457
Gramann M., 1993, ApJ, 405, L47
Granett B. R., Neyrinck M. C., Szapudi I., 2008, ApJ, 683, L99
Granett B. R., et al., 2015, A&A, 583, A61
Gunn J. E., et al., 2006, AJ, 131, 2332
Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111
Guo H., et al., 2015a, MNRAS, 446, 578
Guo H., et al., 2015b, MNRAS, 453, 4368
Guo H., et al., 2016, preprint, (arXiv:1508.07012)
Guzzo L., et al., 2008, Nature, 451, 541
Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489
Hahn O., Angulo R. E., Abel T., 2015, MNRAS, 454, 3920
Hamilton A. J. S., Tegmark M., 2004, MNRAS, 349, 115
Heitmann K., White M., Wagner C., Habib S., Higdon D., 2010, ApJ, 715, 104
Hernández-Monteagudo C., Ma Y.-Z., Kitaura F.-S., Wang W., Génova-Santos R., Macías-Pérez J., Herranz D., 2015, Physical Review Letters, 115, 191301
Heß S., Kitaura F.-S., 2016, MNRAS, 456, 4247
Heß S., Kitaura F.-S., Gottlöber S., 2013, MNRAS, 435, 2065
Ilić S., Langer M., Douspis M., 2013, A&A, 556, A51
Jain B., Zhang P., 2008, Phys. Rev. D, 78, 063503
Jasche J., Kitaura F.-S., 2010, MNRAS, 407, 29
Jasche J., Wandelt B. D., 2013, MNRAS, 432, 894
Jasche J., Kitaura F.-S., Li C., Enßlin T. A., 2010, MNRAS, pp 1638–+
Jasche J., Leclercq F., Wandelt B. D., 2015, J. Cosmology Astropart. Phys., 1, 036
Jennings E., 2012, MNRAS, 427, L25
Jennings E., Jennings D., 2015, MNRAS, 449, 3407
Jennings E., Baugh C. M., Pascoli S., 2011, MNRAS, 410, 2081
Joachimi B., Taylor A. N., Kiessling A., 2011, MNRAS, 418, 145
Kaiser N., 1987, MNRAS, 227, 1
Kim J., Park C., Choi Y.-Y., 2008, ApJ, 683, 123
Kitaura F.-S., 2013, MNRAS, 429, L84
Kitaura F.-S., Angulo R. E., 2012, MNRAS, 425, 2443
Kitaura F.-S., Enßlin T. A., 2008, MNRAS, 389, 497
Kitaura F.-S., Heß S., 2013, MNRAS, 435, L78
Kitaura F.-S., Jasche J., Li C., Enßlin T. A., Metcalf R. B., Wandelt B. D., Lemson G., White S. D. M., 2009, MNRAS, 400, 183
Kitaura F.-S., Jasche J., Metcalf R. B., 2010, MNRAS, 403, 589
Kitaura F.-S., Gallerani S., Ferrara A., 2012a, MNRAS, 420, 61
Kitaura F.-S., Angulo R. E., Hoffman Y., Gottlöber S., 2012b, MNRAS, 425, 2422
Kitaura F.-S., Erdogdu P., Nuza S. E., Khalatyan A., Angulo R. E., Hoffman Y., Gottlöber S., 2012c, MNRAS, 427, L35
Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21
Kitaura F.-S., Gil-Marín H., Scóccola C. G., Chuang C.-H., Müller V., Yepes G., Prada F., 2015, MNRAS, 450, 1836
Kitaura F.-S., et al., 2016a, MNRAS, 456, 4156
Kitaura F.-S., Ata M., Angulo R. E., Chuang C.-H., Rodríguez-Torres S., Monteagudo C. H., Prada F., Yepes G., 2016b, MNRAS, 457, L113
Klypin A., Holtzman J., 1997, ArXiv Astrophysics e-prints,
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Kudlicki A., Chodorowski M., Plewa T., Różyczka M., 2000, MNRAS, 316, 464
Kwan J., Lewis G. F., Linder E. V., 2012, ApJ, 748, 78
LSST Dark Energy Science Collaboration 2012, preprint, (arXiv:1211.0310)
Laureijs R., 2009, preprint, (arXiv:0912.0914)

Lavaux G., Mohayaee R., Colombi S., Tully R. B., Bernardeau F., Silk J., 2008, MNRAS, 383, 1292

Lavaux G., Tully R. B., Mohayaee R., Colombi S., 2010, ApJ, 709, 483

Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, ApJ, 738, 45

Leclercq F., Jasche J., Gil-Marín H., Wandelt B., 2013, J. Cosmology Astropart. Phys., 11, 48

McCullagh N., Neyrinck M., Norberg P., Cole S., 2016, MNRAS, 457, 3652

McDonald P., Roy A., 2009, J. Cosmology Astropart. Phys., 8, 20

McDonald P., Seljak U., 2009, J. Cosmology Astropart. Phys., 10, 7

Mohayaee R., Tully R. B., 2005, ApJ, 635, L113

Monaco P., Efstathiou G., 1999, MNRAS, 308, 763

Nadkarni-Ghosh S., Singhal A., 2016, MNRAS, 457, 2773

Neal R. M., 1993, in Technical Report CRG-TR-93-1. Dept. of Computer Science, University of Toronto

Neal R. M., 2012, preprint, (arXiv:1206.1901)

Nesseris S., Perivolaropoulos L., 2008, Phys. Rev. D, 77, 023504

Neyrinck M. C., Hamilton A. J. S., Gnedin N. Y., 2004, MNRAS, 348, 1

Neyrinck M. C., Szapudi I., Szalay A. S., 2009, ApJ, 698, L90

Neyrinck M. C., Szapudi I., Szalay A. S., 2011, ApJ, 731, 116

Neyrinck M. C., Aragón-Calvo M. A., Jeong D., Wang X., 2014, MNRAS, 441, 646

Nusser A., Dekel A., Bertschinger E., Blumenthal G. R., 1991, ApJ, 379, 6

Nusser A., Davis M., Branchini E., 2014, ApJ, 788, 157

Nuza S. E., Sánchez A. G., Prada F., Klypin A., Schlegel D. J., Gottlöber S., Montero-Dorta A. D., et al 2013, MNRAS, 432, 743

Nuza S. E., Kitaura F.-S., Heß S., Libeskind N. I., Müller V., 2014, MNRAS, 445, 988

Okumura T., Matsubara T., Eisenstein D. J., Kayo I., Hikage C., Szalay A. S., Schneider D. P., 2008, ApJ, 676, 889

Okumura T., Seljak U., Desjacques V., 2012, J. Cosmology Astropart. Phys., 11, 14

Okumura T., Seljak U., Vlah Z., Desjacques V., 2014, J. Cosmology Astropart. Phys., 5, 3

Padmanabhan N., Xu X., Eisenstein D. J., Scalzo R., Cuesta A. J., Mehta K. T., Kazin E., 2012, MNRAS, 427, 2132

Peacock J. A., et al., 2001, Nature, 410, 169

Percival W. J., White M., 2009, MNRAS, 393, 297

Percival W. J., et al., 2004, MNRAS, 353, 1201

Planck Collaboration et al., 2016, A&A, 586, A140

Reid B. A., et al., 2012, MNRAS, 426, 2719

Reid B. A., Seo H.-J., Leauthaud A., Tinker J. L., White M., 2014, MNRAS, 444, 476

Reid B., et al., 2016, MNRAS, 455, 1553

Rodríguez-Torres S. A., et al., 2015, preprint, (arXiv:1509.06404)

Ross et al. A. J., 2015, MNRAS, 449, 835

Samushia L., Percival W. J., Raccanelli A., 2012, MNRAS, 420, 2102

Samushia L., et al., 2013, MNRAS, 429, 1514

Samushia L., et al., 2014, MNRAS, 439, 3504

Sánchez A. G., et al., 2014, MNRAS, 440, 2692

Schaan E., et al., 2015, preprint, (arXiv:1510.06442)

Schlegel D., Abdalla F., Abraham T., Ahn C., Allende Prieto C., Annis J., Aubourg E., et al 2011, preprint, (arXiv:1106.1706)

Seljak U., 2012, J. Cosmology Astropart. Phys., 3, 004

Simpson F., et al., 2016, Phys. Rev. D, 93, 023525

Smee S. A., et al., 2013, AJ, 146, 32

Song Y.-S., Koyama K., 2009, J. Cosmology Astropart. Phys., 1, 48

Song Y.-S., Percival W. J., 2009, J. Cosmology Astropart. Phys., 10, 4

Song Y.-S., Sabiu C. G., Nichol R. C., Miller C. J., 2010, J. Cosmology Astropart. Phys., 1, 25

Song Y.-S., Sabiu C. G., Kayo I., Nichol R. C., 2011, J. Cosmology Astropart. Phys., 5, 20

Springel V., et al., 2005, Nature, 435, 629

Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, MNRAS, 387, 1391

Tasitsiomi A., Kravtsov A. V., Gottlöber S., Klypin A. A., 2004, ApJ, 607, 125

Taylor J. F., Ashdown M. A. J., Hobson M. P., 2008, MNRAS, 389, 1284

Tojeiro R., et al., 2014, MNRAS, 440, 2222

Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16

Vale A., Ostriker J. P., 2004, MNRAS, 353, 189

Vargas-Magaña M., Ho S., Fromenteau S., Cuesta A. J., 2015, preprint, (arXiv:1509.06384)

Wang Y., 2014, MNRAS, 443, 2950

Wang H., Mo H. J., Yang X., van den Bosch F. C., 2012, MNRAS, 420, 1809

Wang H., Mo H. J., Yang X., van den Bosch F. C., 2013, ApJ, 772, 63

Wang H., Mo H. J., Yang X., Jing Y. P., Lin W. P., 2014, ApJ, 794, 94

Watkins R., Feldman H. A., Hudson M. J., 2009, MNRAS, 392, 743

Wetzel A. R., White M., 2010, MNRAS, 403, 1072

White M., Song Y.-S., Percival W. J., 2009, MNRAS, 397, 1348

White M., Blanton M., Bolton A., Schlegel D., Tinker J., Berlind A., da Costa L., et al 2011, ApJ, 728, 126

White M., Tinker J. L., McBride C. K., 2014, MNRAS, 437, 2594

Yahil A., Strauss M. A., Davis M., Huchra J. P., 1991, ApJ, 372, 380

Yu Y., Zhang P., Lin W., Cui W., Fry J. N., 2011, Phys. Rev. D, 84, 023523

Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, ApJ, 449, 446

Zel'dovich Y. B., 1970, A&A, 5, 84

Zhang P., Liguori M., Bean R., Dodelson S., 2007, Physical Review Letters, 99, 141302

Zhao G.-B., Giannantonio T., Pogosian L., Silvestri A., Bacon D. J., Koyama K., Nichol R. C., Song Y.-S., 2010, Phys. Rev. D, 81, 103510

Zheng Y., Zhang P., Jing Y., Lin W., Pan J., 2013, Phys. Rev. D, 88, 103510

da Ângela J., et al., 2008, MNRAS, 383, 565

de Jong R. S., et al., 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. (arXiv:1206.6885), doi:10.1117/12.926239

de la Torre S., Peacock J. A., 2013, MNRAS, 435, 743

de la Torre S., et al., 2013, A&A, 557, A54

[1] Leibniz Institut für Astrophysik (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany
[2] Lawrence Berkeley National Lab, 1 Cyclotron Rd, Berkeley CA 94720, USA
[3] Departments of Physics and Astronomy, University of California, Berkeley, CA 94720, USA
[4] Instituto de Astrofisica de Canarias (IAC), Calle Via Lactea s/n, 38200
[5] Departamento de Astrofísica, Universidad de La Laguna (ULL), E-38206 La Laguna, Tenerife, Spain
[6] Instituto de Física Teórica, (UAM/CSIC), Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain

[7]Campus of International Excellence UAM+CSIC, Cantoblanco, E-28049 Madrid, Spain

[8]Departamento de Física Teórica, Universidad Autónoma de Madrid, Cantoblanco, 28049, Madrid, Spain

[9]Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Plaza San Juan, 1, planta 2, E-44001 Teruel, Spain,

[10] Sorbonne Universités, Institut Lagrange de Paris (ILP), 98 bis Boulevard Arago, 75014 Paris, France

[11] Laboratoire de Physique Nucléaire et de Hautes Energies, Université Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris, France

[12] Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK

[13] Department of Physics and Astronomy, University of Utah, 115 S 1400 E, Salt Lake City, UT 84112, USA

[14] Yale Center for Astronomy and Astrophysics, Yale University, New Haven, CT, 06520, USA

[15] Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA

[16] Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Shanghai 200030, China

[17] Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

[18] Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA

[19] Department of Chemistry and Physics, King's College, 133 North River St, Wilkes Barre, PA 18711, USA

[20]Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain

[21] Department of Astronomy and Space Science, Sejong University, Seoul, 143-747, Korea

[22] Max-Planck-Institut für extraterrestrische Physik, Postfach 1312, Giessenbachstr., D-85741 Garching, Germany

[23] Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802

[24] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802

[25]Department of Physics and Astronomy, Ohio University, 251B Clippinger Labs, Athens, OH 45701, USA
- La Laguna, Tenerife, Spain

[26]Center for Cosmology and Particle Physics, New York University, 4 Washington Place, New York, NY 10003, USA

[27] University of St Andrews, North Haugh, St Andrews Fife, KY16 9SS, UK

[28] Instituto de Física, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, 01000 México, D.F., México

APPENDIX A: COSMIC DENSITY AND VELOCITY RECONSTRUCTION BASED ON THE FULL CMASS BOSS DR12 SAMPLE

Here we present results from the Bayesian reconstruction of cosmic density and peculiar velocity fields applied to the full volume covered by CMASS BOSS DR12 data. The chosen resolution is $6.25\,h^{-1}$ Mpc on a 3D cubical mesh of side $3200\ h^{-1}$ Mpc and 512^3 cells. A random velocity dispersion term was included in the iterative procedure, as explained in §2.2. The completeness, galaxy number counts, and r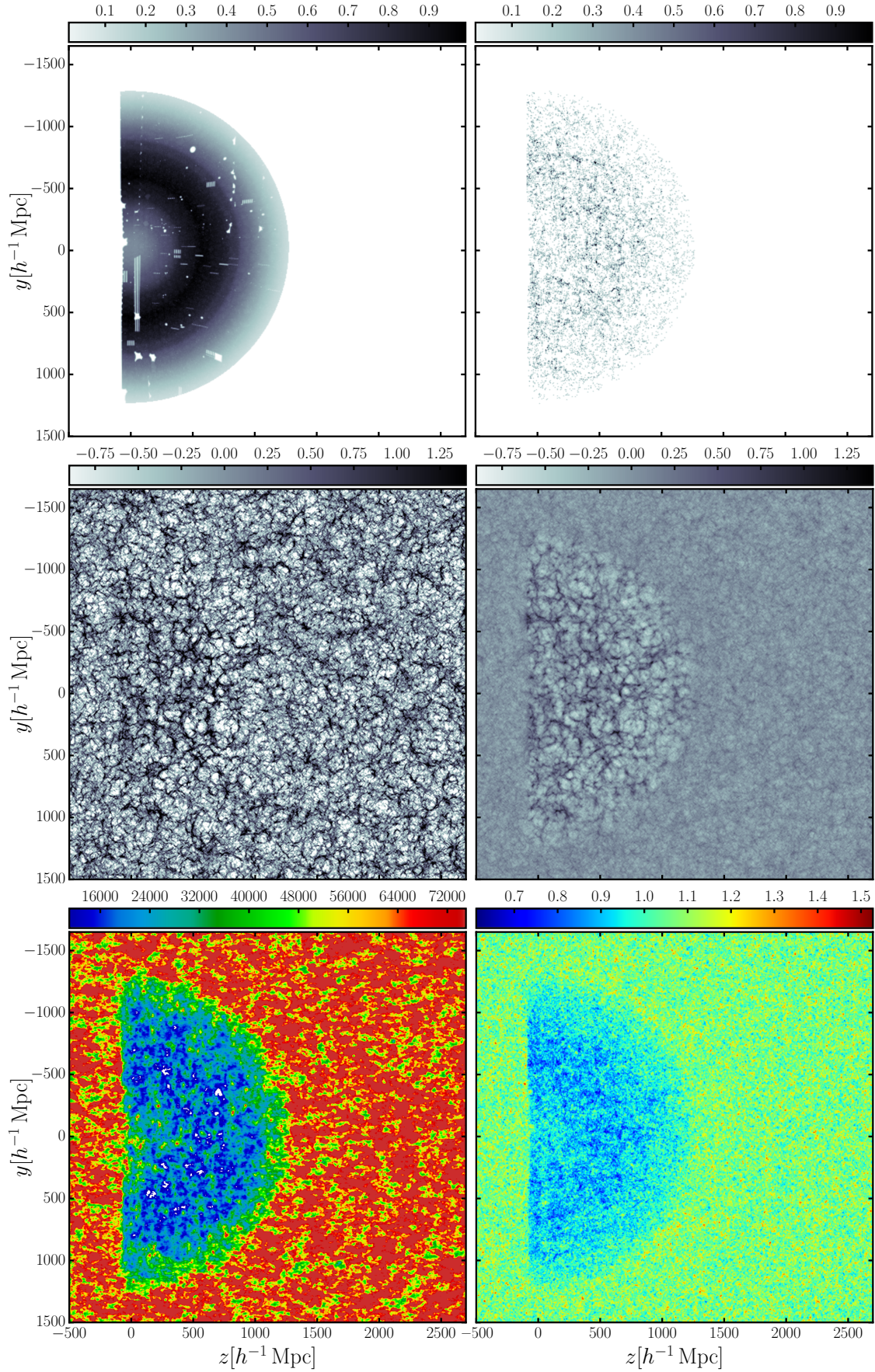econstructions are shown in Fig. A1. The variance of the density and peculiar velocity field reconstructions in the bottom panels show that the uncertainty in the reconstructed fields increase towards lower completeness, being largest in the unobserved regions. The analysis of the reconstructed dark matter density field power spectrum shown in the upper panels of Fig. A2 demonstrates for both mock and observation based catalogues that the reconstructed dark matter fields are unbiased towards large scales ($k \lesssim 0.2\,h\,\mathrm{Mpc}^{-1}$). The correlation function study based on the galaxy catalogues depicted in the middle panels qualitatively demonstrates that we recover the real space correlation function including the real space baryon acoustic peak (see left panel based on mock data). The recovered quadrupole shown in the lower panel has the same features as the original real space mock galaxy catalogue on large scales. These results show that the method presented in this work is handling correctly the selection effects, biasing, and peculiar motions on large scales.

Figure A1. Based on the CMASS BOSS DR12 catalogue with $6.25\,h^{-1}$ Mpc resolution: slices of the 3D cubical mesh of side 3200 h^{-1} Mpc and 512^3 cells. **Upper left panel:** the 3D completeness or window function function, **upper right panel:** galaxy number count per cell. **Middle left panel:** one reconstructed Zel'dovich transformed density field, **middle right panel:** mean over 6000 reconstructed Zel'dovich transformed density fields. **Lower left panel:** corresponding variance of the peculiar velocity field, **lower right panel:** corresponding variance of the Zel'dovich transformed density field.
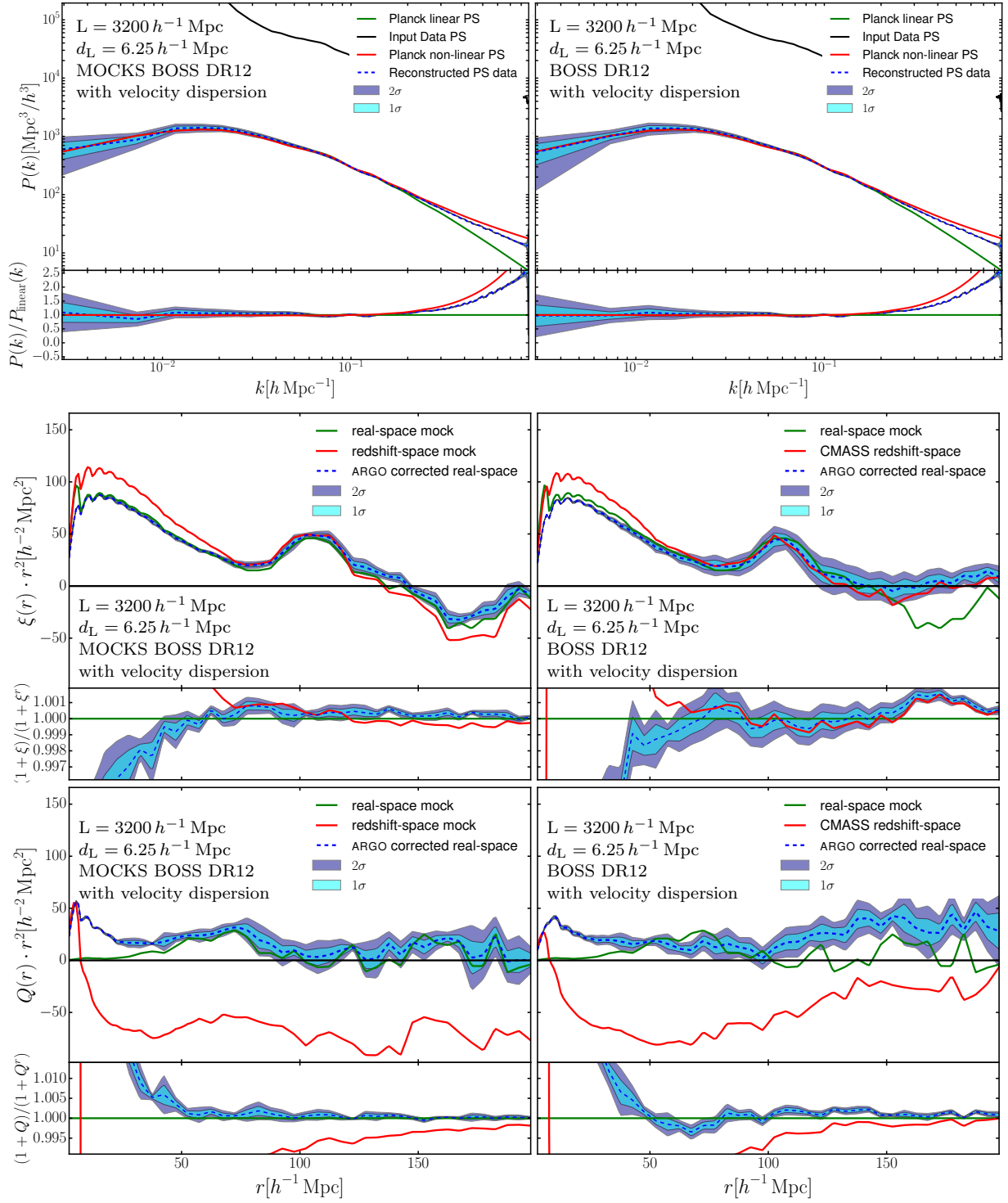
Figure A2. Monopoles and quadrupoles based on the (**left panels:**) light-cone mock catalogue, (**right panels:**) CMASS BOSS DR12 data. Same setting as in previous figure. **Upper panels:** power spectra showing the mean (dashed blue line) over 6000 reconstructed dark matter fields on a mesh with 1 and 2 σ contours (light and dark blue shaded areas, respectively), as compared to the raw galaxy power spectrum (black solid line), the nonlinear (red solid line) and the linear power spectrum (green solid line) assuming the fiducial cosmology, and below the corresponding ratio with respect to the linear power spectrum. **Middle panels:** two-point correlation functions of the galaxy distribution showing the mean (dashed blue line) over 6000 reconstructed real space catalogues with 1 and 2 σ contours (light and dark blue shaded areas, respectively), in addition, the real (green line for mocks only) and redshift space (red line) catalogues, and below the corresponding ratio with respect to the real space correlation function (ξ^{r}). **Lower panels:** corresponding quadrupole correlation functions to the middle panels.